U.S. DEPARTMENT OF COMMERCE
Patent and Trademark Office

SEARCH REQUEST FORM

1.737

Requestor's Name: PHILLIP GAMBEL

Serial Number: 08/487283

Date: 1/28/97

Phone: 308-3997

Art Unit: 1806

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevent citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevent claim(s).

SEQ / INTERFERENCE SEARCH

SEQ ID NO. 1 + 2

REQUEST JAN

Thanks

1/28/98

STAFF USE ONLY

Date completed:	1/28/98	**Search Site**		**Vendors**	
Searcher:			STIC	✓	IG mP
Terminal time:	10	✓	CM-1		STN
Elapsed time:			Pre-S		Dialog
CPU time:		**Type of Search**			APS
Total time:	10		N.A. Sequence		Geninfo
Number of Searches:	1	✓	A.A. Sequence x2		SDC
Number of Databases:	5		Structure		DARC/Questel
			Bibliographic		Other

PTO-1590 (9-90)

****

MPSRCH (TM)

****

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Jan 28 12:10:17 1998; MasPar time 3.57 Seconds
179.105 Million cell updates/sec

Tabular output not generated.

Title: >US-08-487-283A-1
Description: (1-21) from US08487283A.pep
Perfect Score: 141
Sequence: 1 VIDHQGTKSSKCVRQKVEGSS 21

Scoring table: PAM 150
Gap 15

Searched: 95051 seqs, 30469580 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir53
1:ann1 2:ann2 3:ann3 4:ann4 5:unann1 6:unann2 7:unann3
8:unann4 9:unann5 10:unann6 11:unann7 12:unann8
13:unann9 14:unann10 15:unann11 16:unann12 17:unenc
18:unrev

Statistics: Mean 25.513; Variance 35.018; scale 0.729

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	141	100.0	1676	2	C5HU	complement C5 precur	3.46e-20
2	69	48.9	1680	2	C5MS	complement C5 precur	9.17e-03
3	62	44.0	213	11	S76070	hypothetical protein	2.43e-01
4	59	41.8	940	10	S49087	lactoferrin binding	9.30e-01
5	59	41.8	1002	15	A56678	yemanuclein-alpha -	9.30e-01
6	58	41.1	205	16	I49364	protein tyrosine pho	1.44e+00
7	58	41.1	223	16	I49365	protein tyrosine pho	1.44e+00
8	57	40.4	117	8	I68524	ribosomal protein L3	2.22e+00
9	57	40.4	264	18	S28969	N-carbamoylsarcosine	2.22e+00
10	56	39.7	537	13	B334P5	spore coat protein S	3.42e+00
11	56	39.7	540	13	S21825	vicilin-like storage	3.42e+00
12	56	39.7	573	13	A53234	globulin-1S, GLB1S -	3.42e+00
13	55	39.0	775	15	A32494	transposable element	5.22e+00
14	54	38.3	148	12	E47119	spore coat peptide C	7.93e+00
15	54	38.3	224	12	G64383	riboflavin-specific	7.93e+00
16	54	38.3	562	18	S46281	P element - fruit fl	7.93e+00
17	54	38.3	818	1	A30107	dipeptidyl aminopept	7.93e+00
18	54	38.3	1030	14	S57380	probable membrane pr	7.93e+00
19	53	37.6	116	4	Q4ADE5	early E4 11K protein	1.20e+01
20	53	37.6	116	4	Q4ADE2	early E4 11K protein	1.20e+01
21	53	37.6	148	7	S41772	glycine-rich RNA-bin	1.20e+01
22	53	37.6	589	8	C38128	epithelin/granulin p	1.20e+01
23	53	37.6	591	14	S43428	omega-crystallin - g	1.20e+01
24	53	37.6	669	15	I54205	galactosylceramidase	1.20e+01
25	53	37.6	1089	1	PFHUGA	platelet-derived gro	1.20e+01
26	53	37.6	1144	18	A54810	TMV resistance prote	1.20e+01
27	53	37.6	2471	16	A49128	cell-fate determinin	1.20e+01
28	52	36.9	122	13	S21823	Glb1-0 protein - mai	1.80e+01
29	52	36.9	122	13	C53234	globulin-10, GLB10 -	1.80e+01
30	52	36.9	329	3	VMUT21	VSG expression site-	1.80e+01
31	52	36.9	354	2	DUAGO	ornithine cyclodeami	1.80e+01
32	52	36.9	354	6	S55582	methylaspartate ammo	1.80e+01
33	52	36.9	582	13	B53234	vicilin-like storage	1.80e+01
34	52	36.9	982	15	S00954	pol polyprotein - fr	1.80e+01
35	51	36.2	238	9	S23314	hypothetical protein	2.69e+01
36	51	36.2	250	9	S07237	hypothetical protein	2.69e+01
37	51	36.2	337	5	A35260	alcohol dehydrogenas	2.69e+01
38	51	36.2	462	5	S29004	pyruvate kinase (EC	2.69e+01
39	51	36.2	529	4	VGNVAC	major envelope glyco	2.69e+01
40	51	36.2	574	8	A37197	estrogen receptor -	2.69e+01
41	51	36.2	1084	9	S23319	hypothetical protein	2.69e+01
42	51	36.2	1129	16	A47511	protein-tyrosine kin	2.69e+01
43	51	36.2	1132	16	JC4127	protein-tyrosine kin	2.69e+01
44	51	36.2	1291	9	S05465	retrovirus-related p	2.69e+01
45	51	36.2	1678	8	S52588	clathrin heavy chain	2.69e+01

ALIGNMENTS

RESULT 1
ENTRY C5HU #type complete
TITLE complement C5 precursor - human
CONTAINS C5a anaphylatoxin; C5b
ORGANISM #formal_name Homo sapiens #common_name man
DATE 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 27-Feb-1997
ACCESSIONS A40075; A27689; A01267; A01266; S15121
REFERENCE A40075
#authors Haviland, D.L.; Haviland, J.C.; Fleischer, D.T.; Hunt, A.; Wetsel, R.A.
#journal J. Immunol. (1991) 146:362-368
#title Complete cDNA sequence of human complement pro-C5. Evidence of truncated transcripts derived from a single copy gene.
#cross-references MUID:91079575
#accession A40075
##molecule_type mRNA
##residues 1-1676 ##label HAV
##cross-references GB:M57729
##note 518-Ser was also found
REFERENCE A27689
#authors Wetsel, R.A.; Lemons, R.S.; Le Beau, M.M.; Barnum, S.R.; Noack, D.; Tack, B.F.
#journal Biochemistry (1988) 27:1474-1482
#title Molecular analysis of human complement component C5: localization of the structural gene to chromosome 9.

```
   #cross-references MUID:88209511
   #accession     A27689
      ##molecule_type mRNA
      ##residues      412-1676 ##label WET
      ##cross-references GB:M18879
REFERENCE        A01267
   #authors      Fernandez, H.N.; Hugli, T.E.
   #journal      J. Biol. Chem. (1978) 253:6955-6964
   #title        Primary structural analysis of the polypeptide portion of
                   human C5a anaphylatoxin. Polypeptide sequence determination
                   and assignment of the oligosaccharide attachment site in
                   C5a.
   #cross-references MUID:79005687
   #accession     A01267
      ##molecule_type protein
      ##residues      678-751 ##label FER
REFERENCE        A01266
   #authors      Lundwall, A.B.; Wetsel, R.A.; Kristensen, T.; Whitehead,
                   A.S.; Woods, D.E.; Ogden, R.C.; Colten, H.R.; Tack, B.F.
   #journal      J. Biol. Chem. (1985) 260:2108-2112
   #title        Isolation and sequence analysis of a cDNA clone encoding the
                   fifth complement component.
   #cross-references MUID:85130937
   #accession     A01266
      ##molecule_type mRNA
      ##residues      412-854,
                        'SLALSPRLECNGKISGHCKLRLPGSSDSPASASQVAGITGTHHHAQPT'
                        ##label LUN
      ##cross-references GB:K02874
      ##note          the carboxyl-terminal part of the sequence in this
                        report appears to be derived from translation of an
                        ALU repeat sequence
REFERENCE        S15121
   #authors      Bohnsack, J.F.; Mollison, K.W.; Buko, A.M.; Ashworth, J.C.;
                   Hill, H.R.
   #journal      Biochem. J. (1991) 273:635-640
   #title        Group B streptococci inactivate complement component C5a by
                   enzymic cleavage at the C-terminus.
   #cross-references MUID:91144547
   #contents     annotation
COMMENT    Complement C5 contains two disulfide-linked chains, formed by
             removal of four basic residues. C5 convertase releases C5a
             anaphylatoxin from the amino end of the alpha chain, generating
             C5b (beta and alpha' chains).
COMMENT    Activation of C5 initiates the spontaneous assembly of the late
             complement components, C5-C9, into the membrane attack complex.
             C5b has a transient binding site for C6. The C5b-C6 complex is
             the foundation upon which the membrane attack complex is
             assembled.
COMMENT    C5a has potent spasmogenic and chemotactic activity.
GENETICS
   #gene         GDB:C5
      ##cross-references GDB:119734
   #map_position 9q33-9q33
CLASSIFICATION   #superfamily alpha-2-macroglobulin
KEYWORDS         complement alternate pathway; complement pathway; cytolysis;
                   glycoprotein; inflammation; membrane attack complex; plasma
FEATURE
   1-18               #domain signal sequence #status predicted #label SIG\
   19-673,678-1676    #product complement C5 #status predicted #label MAT\
   19-673,752-1676    #product C5b #status predicted #label C5B\
   19-673             #product complement C5 and C5b beta chain #status
                        predicted #label C5BB\
   678-1676           #product complement C5 alpha chain #status predicted
```



```
                        #label C5A\
   678-751              #product C5a anaphylatoxin #status experimental #label
                          C5T\
   752-1676             #product C5b alpha' chain #status predicted #label C5BA\
   567-810,634-669,
   698-724,699-731,
   711-732,866-1527,
   1101-1159,
   1375-1505,
   1405-1474,
   1520-1525,
   1532-1606,
   1553-1676,
   1654-1657            #disulfide_bonds #status predicted\
   741                  #binding_site carbohydrate (Asn) (covalent) #status
                          experimental\
   751-752              #cleavage_site Arg-Leu (C5 convertase) #status
                          experimental\
   911,1115,1630        #binding_site carbohydrate (Asn) (covalent) #status
                          predicted
SUMMARY          #length 1676  #molecular-weight 188330  #checksum 3858

  Query Match          100.0%;  Score 141;  DB 2;  Length 1676;
  Best Local Similarity 100.0%;  Pred. No. 3.46e-20;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Db     872 vidhqgtksskcvrqkvegss 892
           |||||||||||||||||||||
Qy       1 VIDHQGTKSSKCVRQKVEGSS 21


RESULT    2
ENTRY            C5MS       #type complete
TITLE            complement C5 precursor - mouse
CONTAINS         C5a anaphylatoxin; C5b
ORGANISM         #formal_name Mus musculus #common_name house mouse
DATE             19-Nov-1988 #sequence_revision 15-Oct-1994 #text_change
                   16-Feb-1997
ACCESSIONS       A35530; A27538; A40429
REFERENCE        A35530
   #authors      Wetsel, R.A.; Fleischer, D.T.; Haviland, D.L.
   #journal      J. Biol. Chem. (1990) 265:2435-2440
   #title        Deficiency of the murine fifth complement component (C5). A
                   2-base pair gene deletion in a 5'-exon.
   #cross-references MUID:90153853
   #accession     A35530
      ##molecule_type mRNA
      ##residues      1-215,'L' ##label WET
      ##cross-references GB:J05234
REFERENCE        A27538
   #authors      Wetsel, R.A.; Ogata, R.T.; Tack, B.F.
   #journal      Biochemistry (1987) 26:737-743
   #title        Primary structure of the fifth component of murine
                   complement.
   #cross-references MUID:87185363
   #accession     A27538
      ##molecule_type mRNA
      ##residues      'PGL',44-1680 ##label WET2
REFERENCE        A40429
   #authors      Haviland, D.L.; Haviland, J.C.; Fleischer, D.T.; Wetsel, R.A.
   #journal      J. Biol. Chem. (1991) 266:11818-11825
   #title        Structure of the murine fifth complement component (C5) gene.
                   A large, highly interrupted gene with a variant donor
                   splice site and organizational homology with the third and
```

```
                 fourth complement component genes.
   #cross-references MUID:91268053
   #accession    A40429
      ##molecule_type DNA
      ##residues      1-15 ##label HAV
      ##cross-references GB:M64852
COMMENT    Complement C5 contains two disulfide-linked chains, formed by
             removal of four basic residues. C5 convertase releases C5a
             anaphylatoxin from the amino end of the alpha chain, generating
             C5b (beta and alpha' chains).
COMMENT    Activation of C5 initiates the spontaneous assembly of the late
             complement components, C5-C9, into the membrane attack complex.
             C5b has a transient binding site for C6. The C5b-C6 complex is
             the foundation upon which the membrane attack complex is
             assembled.
COMMENT    C5a has potent spasmogenic and chemotactic activity.
GENETICS
   #map_position 2
   #introns       22/3; 86/3; 140/3; 164/3; 195/2; 223/1; 253/2; 291/3; 334/1;
                    372/3; 434/3; 502/3; 572/3; 622/3; 667/1; 691/1; 757/1;
                    787/2; 812/1; 858/3; 934/3; 955/1; 985/1; 1056/1; 1081/2;
                    1134/3; 1166/3; 1224/1; 1292/3; 1343/3; 1364/3; 1392/1;
                    1411/2; 1445/3; 1470/3; 1506/1; 1534/1; 1564/1; 1592/1;
                    1637/2
CLASSIFICATION   #superfamily alpha-2-macroglobulin
KEYWORDS         complement alternate pathway; complement pathway; cytolysis;
                   glycoprotein; inflammation; membrane attack complex; plasma
FEATURE
   1-18               #domain signal sequence #status predicted #label SIG\
   19-674,679-1679    #product complement C5 #status predicted #label MAT\
   19-674,756-1679    #product C5b #status predicted #label C5B\
   19-674             #product complement C5 and C5b beta chain #status
                        predicted #label C5BB\
   679-1679           #product complement C5 alpha chain #status predicted
                        #label C5A\
   679-755            #product C5a anaphylatoxin #status predicted #label C5T\
   756-1679           #product C5b alpha' chain #status predicted #label C5BA\
   567-814,635-670,
   702-728,703-735,
   715-736,870-1531,
   1105-1163,
   1379-1509,
   1409-1478,
   1524-1529,
   1536-1609,
   1557-1679,
   1657-1660          #disulfide_bonds #status predicted\
   915,1119,1633      #binding_site carbohydrate (Asn) (covalent) #status
                        predicted
SUMMARY          #length 1680  #molecular-weight 188876  #checksum 3888

  Query Match            48.9%;  Score 69;  DB 2;  Length 1680;
  Best Local Similarity  47.1%;  Pred. No. 9.17e-03;
  Matches    8;  Conservative    7;  Mismatches   2;  Indels    0;  Gaps    0;

Db     880 htsrpsrcvfqriegss 896
           : :::|:|| |::||||
Qy       5 QGTKSSKCVRQKVEGSS 21


RESULT    3
ENTRY            S76070     #type complete
TITLE            hypothetical protein - Synechocystis sp. (PCC 6803)
ORGANISM         #formal_name Synechocystis sp.
```



```
   #variety      PCC 6803
DATE             25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
                   25-Apr-1997
ACCESSIONS       S76070
REFERENCE        S74322
   #authors      Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
                   Nakamura, Y.; Miyajima, N.; Hirosawa, M.; Sugiura, M.;
                   Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
                   Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
                   S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
                   Yasuda, M.; Tabata, S.
   #journal      DNA Res. (1996) 3:109-136
   #title        Sequence analysis of the genome of the unicellular
                   cyanobacterium Synechocystis sp. PCC6803. II. Sequence
                   determination of the entire genome and assignment of
                   potential protein-coding regions.
   #accession    S76070
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-213 ##label KAN
      ##cross-references EMBL:D63999
      ##note          the nucleotide sequence was submitted to the EMBL Data
                        Library, June 1996
SUMMARY          #length 213  #molecular-weight 22745  #checksum 2191

  Query Match            44.0%;  Score 62;  DB 11;  Length 213;
  Best Local Similarity  35.0%;  Pred. No. 2.43e-01;
  Matches    7;  Conservative    7;  Mismatches   6;  Indels    0;  Gaps    0;

Db      92 ilayigakasdcikpkvvss 111
           ::   |:|:| |:: || :|
Qy       1 VIDHQGTKSSKCVRQKVEGS 20


RESULT    4
ENTRY            S49087     #type complete
TITLE            lactoferrin binding protein - Neisseria meningitidis
ORGANISM         #formal_name Neisseria meningitidis
DATE             16-Feb-1995 #sequence_revision 12-May-1995 #text_change
                   12-May-1995
ACCESSIONS       S49087
REFERENCE        S49087
   #authors      Pettersson, A.M.; Klarenbeek, X.Y.Z.; van Deurzen, X.Y.Z.;
                   Poolman, X.Y.Z.; Tommassen, X.Y.Z.
   #submission   submitted to the EMBL Data Library, June 1994
   #description  Molecular characterization of the structural gene for the
                   lacto-ferrin receptor of the meningococcal strain H44/76.
   #accession    S49087
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-940 ##label PET
      ##cross-references EMBL:X79838
SUMMARY          #length 940  #molecular-weight 105347  #checksum 8194

  Query Match            41.8%;  Score 59;  DB 10;  Length 940;
  Best Local Similarity  50.0%;  Pred. No. 9.30e-01;
  Matches    7;  Conservative    5;  Mismatches   2;  Indels    0;  Gaps    0;

Db     592 rsrkcvprkingsn 605
           :| ||| :|::||:
Qy       8 KSSKCVRQKVEGSS 21


RESULT    5
```

```
ENTRY          A56678     #type complete
TITLE          yemanuclein-alpha - fruit fly (Drosophila melanogaster)
ORGANISM       #formal_name Drosophila melanogaster
DATE           08-Jul-1995 #sequence_revision 03-Aug-1995 #text_change
                 16-Feb-1997
ACCESSIONS     A56678; S22146
REFERENCE      A56678
   #authors    Ait-Ahmed, O.; Bellon, B.; Capri, M.; Joblet, C.;
                 Thomas-Delaage, M.
   #journal    Mech. Dev. (1992) 37:69-80
   #title      The yemanuclein-alpha: a new Drosophila DNA binding protein
                 specific for the oocyte nucleus.
   #accession  A56678
      ##status       preliminary
      ##molecule_type DNA
      ##residues     1-1002 ##label AIE
      ##cross-references GB:X63503
GENETICS
   #gene       FlyBase:yem&agr
      ##cross-references FlyBase:FBgn0005596
   #introns    80/3; 154/3; 428/1 477/2; 557/2
KEYWORDS       DNA binding; oocyte
SUMMARY        #length 1002  #molecular-weight 109310  #checksum 4278

  Query Match           41.8%;  Score 59;  DB 15;  Length 1002;
  Best Local Similarity  50.0%;  Pred. No. 9.30e-01;
  Matches     6;  Conservative     5;  Mismatches   1;  Indels    0;  Gaps    0;

Db     47 tktakcirikld 58
          ||::||:| |::
Qy      7 TKSSKCVRQKVE 18


RESULT    6
ENTRY          I49364     #type complete
TITLE          protein tyrosine phosphatase - mouse
ORGANISM       #formal_name Mus musculus #common_name house mouse
DATE           02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
                 02-Jul-1996
ACCESSIONS     I49364
REFERENCE      I49364
   #authors    Wishart, M.J.; Denu, J.M.; Williams, J.A.; Dixon, J.E.
   #journal    J. Biol. Chem. (1995) 270:26782-26785
   #title      A single mutation converts a novel-phosphotyrosine binding
                 domain into a dual-specificity phosphatase.
   #accession  I49364
      ##status       preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues     1-205 ##label RES
      ##cross-references EMBL:U34973; NID:g1063624; CDS_PID:g1063625
SUMMARY        #length 205  #molecular-weight 23683  #checksum 2745

  Query Match           41.1%;  Score 58;  DB 16;  Length 205;
  Best Local Similarity  25.0%;  Pred. No. 1.44e+00;
  Matches     5;  Conservative     9;  Mismatches   6;  Indels    0;  Gaps    0;

Db     51 ilqkhgithiicirqniean 70
          ::: :|     |:||::|::
Qy      1 VIDHQGTKSSKCVRQKVEGS 20


RESULT    7
ENTRY          I49365     #type complete
TITLE          protein tyrosine phosphatase - mouse
```



```
ORGANISM       #formal_name Mus musculus #common_name house mouse
DATE           02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
                 02-Jul-1996
ACCESSIONS     I49365
REFERENCE      I49364
   #authors    Wishart, M.J.; Denu, J.M.; Williams, J.A.; Dixon, J.E.
   #journal    J. Biol. Chem. (1995) 270:26782-26785
   #title      A single mutation converts a novel-phosphotyrosine binding
                 domain into a dual-specificity phosphatase.
   #accession  I49365
      ##status       preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues     1-223 ##label RES
      ##cross-references EMBL:U34973; NID:g1063624; CDS_PID:g1063626
GENETICS
   #introns    168/3
SUMMARY        #length 223  #molecular-weight 25416  #checksum 359

  Query Match           41.1%;  Score 58;  DB 16;  Length 223;
  Best Local Similarity  25.0%;  Pred. No. 1.44e+00;
  Matches     5;  Conservative     9;  Mismatches   6;  Indels    0;  Gaps    0;

Db     51 ilqkhgithiicirqniean 70
          ::: :|     |:||::|::
Qy      1 VIDHQGTKSSKCVRQKVEGS 20


RESULT    8
ENTRY          I68524     #type complete
TITLE          ribosomal protein L34 - human
ORGANISM       #formal_name Homo sapiens #common_name man
DATE           24-May-1996 #sequence_revision 24-May-1996 #text_change
                 16-Feb-1997
ACCESSIONS     I68524
REFERENCE      I54209
   #authors    Rommens, J.M.; Durocher, F.; McArthur, J.; Tonin, P.;
                 Leblanc, J.
   #journal    Genomics (1995) 28:530-542
   #title      Generation of a transcription map at the HSD17B locus
                 centromeric to BRCA1 at 17q21.
   #accession  I68524
      ##status       preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues     1-117 ##label RES
      ##cross-references GB:L38941; NID:g1008855; CDS_PID:g1008856
CLASSIFICATION #superfamily rat ribosomal protein L34
SUMMARY        #length 117  #molecular-weight 13305  #checksum 4392

  Query Match           40.4%;  Score 57;  DB 8;  Length 117;
  Best Local Similarity  41.7%;  Pred. No. 2.22e+00;
  Matches     5;  Conservative     6;  Mismatches   1;  Indels    0;  Gaps    0;

Db     80 gsmcakcvrdri 91
          |:: :||||:::
Qy      6 GTKSSKCVRQKV 17


RESULT    9
ENTRY          S28969     #type complete
TITLE          N-carbamoylsarcosine amidohydrolase (EC 3.5.1.59) -
                 Arthrobacter sp.
ORGANISM       #formal_name Arthrobacter sp.
DATE           19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
                 19-Mar-1997
```

```
ACCESSIONS      S28969
REFERENCE       S28969
   #authors     Romao, M.J.; Turk, D.; Gomis-Rueth, F.X.; Huber, R.;
                  Schumacher, G.; Moellering, H.; Ruessmann, L.
   #journal     J. Mol. Biol. (1992) 226:1111-1130
   #title       Crystal structure analysis, refinement and enzymatic reaction
                  mechanism of N-carbamoylsarcosine amidohydrolase from
                  Arthrobacter sp. at 2.0 A resolution.
   #accession   S28969
      ##status       preliminary
      ##residues     1-264 ##label ROM
SUMMARY         #length 264  #molecular-weight 29057  #checksum 6729

  Query Match            40.4%;  Score 57;  DB 18;  Length 264;
  Best Local Similarity  40.0%;  Pred. No. 2.22e+00;
  Matches    6;  Conservative    5;  Mismatches    4;  Indels    0;  Gaps    0;

Db    171 gataagcvrhtveda 185
          |: :: |||: || :
Qy      6 GTKSSKCVRQKVEGS 20


RESULT   10
ENTRY           B33485     #type complete
TITLE           spore coat protein SP70 - slime mold (Dictyostelium
                  discoideum)
ORGANISM        #formal_name Dictyostelium discoideum
DATE            09-Mar-1990 #sequence_revision 11-Sep-1992 #text_change
                  30-Sep-1993
ACCESSIONS      B33485
REFERENCE       A33485
   #authors     Fosnaugh, K.L.; Loomis, W.F.
   #journal     Mol. Cell. Biol. (1989) 9:5215-5218
   #title       Spore coat genes SP60 and SP70 of Dictyostelium discoideum.
   #cross-references MUID:90097939
   #accession   B33485
      ##status       preliminary
      ##molecule_type DNA; mRNA
      ##residues     1-537 ##label FOS
      ##cross-references GB:M26238
      ##note         the authors translated the codon AAT for residue 281 as
                       Asp
CLASSIFICATION  #superfamily LDL receptor ligand-binding repeat homology
SUMMARY         #length 537  #molecular-weight 56650  #checksum 2250

  Query Match            39.7%;  Score 56;  DB 13;  Length 537;
  Best Local Similarity  54.5%;  Pred. No. 3.42e+00;
  Matches    6;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Db    291 kngecirdkve 301
          |:: |:|:|||
Qy      8 KSSKCVRQKVE 18


RESULT   11
ENTRY           S21825     #type complete
TITLE           vicilin-like storage protein Glb1-S, embryo - maize
ORGANISM        #formal_name Zea mays #common_name maize
DATE            20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
                  20-Feb-1995
ACCESSIONS      S21825
REFERENCE       S21823
   #authors     Kriz, A.L.
   #submission  submitted to the EMBL Data Library, April 1991
```



```
   #accession   S21825
      ##status       preliminary
      ##molecule_type DNA
      ##residues     1-540 ##label KRI
      ##cross-references EMBL:X59084
GENETICS
   #gene        Glb1-S
   #introns     170/1; 195/2; 222/2; 319/2
SUMMARY         #length 540  #molecular-weight 60239  #checksum 1419

  Query Match            39.7%;  Score 56;  DB 13;  Length 540;
  Best Local Similarity  58.3%;  Pred. No. 3.42e+00;
  Matches    7;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Db     32 hgghksgrcvrr 43
          | | ||::|||:
Qy      4 HQGTKSSKCVRQ 15


RESULT   12
ENTRY           A53234     #type complete
TITLE           globulin-1S, GLB1S - maize
ORGANISM        #formal_name Zea mays #common_name maize
DATE            02-May-1994 #sequence_revision 18-Nov-1994 #text_change
                  05-Apr-1995
ACCESSIONS      A53234
REFERENCE       A53234
   #authors     Belanger, F.C.; Kriz, A.L.
   #journal     Genetics (1991) 129:863-872
   #title       Molecular basis for allelic polymorphism of the maize
                  Globulin-1 gene.
   #cross-references MUID:92090707
   #accession   A53234
      ##status       preliminary
      ##molecule_type DNA
      ##residues     1-573 ##label BEL
      ##cross-references NCBIN:71280; NCBIP:71284
      ##experimental_source inbred line Va 26
      ##note         sequence extracted from NCBI backbone
SUMMARY         #length 573  #molecular-weight 65075  #checksum 3569

  Query Match            39.7%;  Score 56;  DB 13;  Length 573;
  Best Local Similarity  58.3%;  Pred. No. 3.42e+00;
  Matches    7;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Db     32 hgghksgrcvrr 43
          | | ||::|||:
Qy      4 HQGTKSSKCVRQ 15


RESULT   13
ENTRY           A32494     #type complete
TITLE           transposable element Txlc protein 1 - African clawed frog
ORGANISM        #formal_name Xenopus laevis #common_name African clawed frog
DATE            12-Oct-1989 #sequence_revision 31-Dec-1993 #text_change
                  31-Dec-1993
ACCESSIONS      A32494
REFERENCE       A32494
   #authors     Garrett, J.E.; Knutzon, D.S.; Carroll, D.
   #journal     Mol. Cell. Biol. (1989) 9:3018-3027
   #title       Composite transposable elements in the Xenopus laevis genome.
   #cross-references MUID:89384562
   #accession   A32494
      ##status       preliminary
```

```
      ##molecule_type DNA
      ##residues      1-775 ##label GAR
      ##cross-references GB:M26915
      ##note          the authors translated the codon ATT for residue as Gln,
                        and AAG for residue 288 as Leu
SUMMARY          #length 775  #molecular-weight 82355  #checksum 6734

  Query Match            39.0%;  Score 55;  DB 15;  Length 775;
  Best Local Similarity  46.7%;  Pred. No. 5.22e+00;
  Matches     7;  Conservative     5;  Mismatches    3;  Indels    0;  Gaps    0;

Db    617 sntskcvssevegtp 631
          :::||||   |||::
Qy      7 TKSSKCVRQKVEGSS 21


RESULT   14
ENTRY            E47119     #type complete
TITLE            spore coat peptide CotZ - Bacillus subtilis
ORGANISM         #formal_name Bacillus subtilis
DATE             21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change
                   18-Nov-1994
ACCESSIONS       E47119
REFERENCE        A47119
   #authors      Zhang, J.; Fitz-James, P.C.; Aronson, A.I.
   #journal      J. Bacteriol. (1993) 175:3757-3766
   #title        Cloning and characterization of a cluster of genes encoding
                   polypeptides present in the insoluble fraction of the spore
                   coat of Bacillus subtilis.
   #cross-references MUID:93285989
   #accession    E47119
      ##status        preliminary
      ##molecule_type nucleic acid
      ##residues      1-148 ##label ZHA
      ##cross-references NCBIN:133538; NCBIP:133548
      ##note          sequence extracted from NCBI backbone
SUMMARY          #length 148  #molecular-weight 16534  #checksum 4681

  Query Match            38.3%;  Score 54;  DB 12;  Length 148;
  Best Local Similarity  63.6%;  Pred. No. 7.93e+00;
  Matches     7;  Conservative     2;  Mismatches    2;  Indels    0;  Gaps    0;

Db      4 ktsscvreave 14
          |:| |||: ||
Qy      8 KSSKCVRQKVE 18


RESULT   15
ENTRY            G64383     #type complete
TITLE            riboflavin-specific deaminase (EC 3.5.4.-) - Methanococcus
                   jannaschii
ORGANISM         #formal_name Methanococcus jannaschii
DATE             13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
                   13-Sep-1996
ACCESSIONS       G64383
REFERENCE        A64300
   #authors      Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann,
                   R.D.; Sutton, G.G.; Blake, J.A.; FitzGerald, L.M.; Clayton,
                   R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.;
                   Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.;
                   Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
                   Scott, J.L.; Geoghagen, N.S.M.; Weidman, J.F.; Fuhrmann,
                   J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson,
                   J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts,
```



```
                   K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk,
                   H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
   #journal      Science (1996) 273:1058-1073
   #title        Complete genome sequence of the methanogenic archaeon,
                   Methanococcus jannaschii.
   #accession    G64383
      ##status        preliminary; nucleic acid sequence not shown;
                        translation not shown
      ##molecule_type DNA
      ##residues      1-224 ##label BUL
      ##cross-references GB:L77117; TIGR:MJ0671; CDS_PID:g1510756
GENETICS
   #map_position REV597638-596964
   #start_codon  TTG
KEYWORDS         hydrolase
SUMMARY          #length 224  #molecular-weight 25037  #checksum 2215

  Query Match            38.3%;  Score 54;  DB 12;  Length 224;
  Best Local Similarity  33.3%;  Pred. No. 7.93e+00;
  Matches     6;  Conservative     4;  Mismatches    8;  Indels    0;  Gaps    0;

Db    118 iledmgvevvkcgrgkvd 135
            :::  |     || | ||:
Qy      1 VIDHQGTKSSKCVRQKVE 18


Search completed: Wed Jan 28 12:10:33 1998
Job time : 16 secs.
```

```
******************************************************************************
MPSRCH (TM)
******************************************************************************
```

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Jan 28 12:10:52 1998; MasPar time 2.26 Seconds
197.180 Million cell updates/sec

Tabular output not generated.

Title: >US-08-487-283A-1
Description: (1-21) from US08487283A.pep
Perfect Score: 141
Sequence: 1 VIDHQGTKSSKCVRQKVEGSS 21

Scoring table: PAM 150
Gap 15

Searched: 59021 seqs, 21210388 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot34
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11

Statistics: Mean 26.425; Variance 29.915; scale 0.883

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	141	100.0	1676	2	CO5_HUMAN	COMPLEMENT C5 PRECURS	1.16e-24
2	69	48.9	1680	2	CO5_MOUSE	COMPLEMENT C5 PRECURS	7.22e-04
3	59	41.8	943	5	IROA_NEIME	IRON-REGULATED OUTER	1.65e-01
4	59	41.8	1002	11	YEMA_DROME	YEMANUCLEIN-ALPHA.	1.65e-01
5	57	40.4	116	8	RL34_HUMAN	60S RIBOSOMAL PROTEIN	4.57e-01
6	57	40.4	264	2	CSH_ARTSP	N-CARBAMOYLSARCOSINE	4.57e-01
7	56	39.7	537	9	SP70_DICDI	SPORE COAT PROTEIN SP	7.54e-01
8	56	39.7	573	4	GLB1_MAIZE	GLOBULIN-1 S ALLELE P	7.54e-01
9	56	39.7	1357	8	RPOB_PSEPU	DNA-DIRECTED RNA POLY	7.54e-01
10	55	39.0	775	11	YTX1_XENLA	TRANSPOSON TX1 HYPOTH	1.24e+00
11	54	38.3	148	2	COTZ_BACSU	SPORE COAT PROTEIN Z.	2.01e+00
12	54	38.3	818	3	DAP2_YEAST	DIPEPTIDYL AMINOPEPTI	2.01e+00
13	54	38.3	1039	11	YR71_CAEEL	HYPOTHETICAL 118.2 KD	2.01e+00
14	53	37.6	116	3	E411_ADE05	PROBABLE EARLY E4 11	3.25e+00
15	53	37.6	116	3	E411_ADE02	PROBABLE EARLY E4 11	3.25e+00
16	53	37.6	589	4	GRN_MOUSE	GRANULINS PRECURSOR (	3.25e+00
17	53	37.6	669	4	GALC_HUMAN	GALACTOCEREBROSIDASE	3.25e+00
18	53	37.6	1089	7	PGDS_HUMAN	ALPHA PLATELET-DERIVE	3.25e+00
19	52	36.9	329	3	ESG2_TRYBB	VSG EXPRESSION SITE-A	5.22e+00
20	52	36.9	354	7	OCD_AGRT5	ORNITHINE CYCLODEAMIN	5.22e+00
21	52	36.9	455	7	P2X5_RAT	P2X PURINOCEPTOR 5 (A	5.22e+00
22	51	36.2	121	11	YIF2_YEAST	PROBABLE 60S RIBOSOMA	8.32e+00
23	51	36.2	337	10	VP67_NPVGM	MAJOR ENVELOPE GLYCOP	8.32e+00
24	51	36.2	337	1	ADH1_ZYMMO	ALCOHOL DEHYDROGENASE	8.32e+00
25	51	36.2	462	5	KPY1_ECOLI	PYRUVATE KINASE I (EC	8.32e+00
26	51	36.2	512	10	VP67_NPVAC	MAJOR ENVELOPE GLYCOP	8.32e+00
27	51	36.2	574	3	ESTR_SALIR	ESTROGEN RECEPTOR (ER	8.32e+00
28	51	36.2	588	4	GRN_RAT	GRANULINS PRECURSOR (	8.32e+00
29	51	36.2	1056	6	MUC5_HUMAN	TRACHEOBRONCHIAL MUCI	8.32e+00
30	51	36.2	1651	10	VIT6_CAEEL	VITELLOGENIN 6 PRECUR	8.32e+00
31	51	36.2	1678	2	CLH_DROME	CLATHRIN HEAVY CHAIN.	8.32e+00
32	50	35.5	123	8	RL14_HAEIN	50S RIBOSOMAL PROTEIN	1.32e+01
33	50	35.5	171	11	YJB6_YEAST	HYPOTHETICAL 19.7 KD	1.32e+01
34	50	35.5	192	10	Y045_NPVAC	HYPOTHETICAL 22.7 KD	1.32e+01
35	50	35.5	296	8	RBSB_ECOLI	D-RIBOSE-BINDING PERI	1.32e+01
36	50	35.5	296	8	RBSB_SALTY	D-RIBOSE-BINDING PERI	1.32e+01
37	50	35.5	301	9	TRT2_CHICK	TROPONIN T, CARDIAC M	1.32e+01
38	50	35.5	410	1	BEDA_PSEPU	BENZENE 1,2-DIOXYGENA	1.32e+01
39	50	35.5	592	5	INVA_DAUCA	BETA-FRUCTOFURANOSIDA	1.32e+01
40	50	35.5	1010	10	WNT5_DROME	PROTEIN DWNT-5 PRECUR	1.32e+01
41	49	34.8	72	10	VSM0_TRYBB	VARIANT SURFACE GLYCO	2.06e+01
42	49	34.8	121	8	RK14_ODOSI	CHLOROPLAST 50S RIBOS	2.06e+01
43	49	34.8	267	11	YD6A_SCHPO	HYPOTHETICAL 30.6 KD	2.06e+01
44	49	34.8	509	10	VP67_NPVCF	MAJOR ENVELOPE GLYCOP	2.06e+01
45	49	34.8	784	4	GCF_HUMAN	GC-RICH SEQUENCE DNA-	2.06e+01

ALIGNMENTS

```
RESULT   1
ID   CO5_HUMAN     STANDARD;     PRT;  1676 AA.
AC   P01031;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   COMPLEMENT C5 PRECURSOR (CONTAINS: C5A ANAPHYLATOXIN).
GN   C5.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91079575.
RA   HAVILAND D.L., HAVILAND J.C., FLEISCHER D.T., HUNT A., WETSEL R.A.;
RL   J. IMMUNOL. 146:362-368(1991).
RN   [2]
RP   SEQUENCE OF 412-1676 FROM N.A.
RX   MEDLINE; 88209511.
RA   WETSEL R.A., LEMONS R.S., LEBEAU M.M., BARNUM S.R., NOACK D.,
RA   TACK B.F.;
RL   BIOCHEMISTRY 27:1474-1482(1988).
RN   [3]
RP   SEQUENCE OF 412-902 FROM N.A.
RX   MEDLINE; 85130937.
RA   LUNDWALL A.B., WETSEL R.A., KRISTENSEN T., WHITEHEAD A.S.,
RA   WOODS D.E., OGDEN R.C., COLTEN H.R., TACK B.F.;
RL   J. BIOL. CHEM. 260:2108-2112(1985).
```

```
RN   [4]
RP   SEQUENCE OF 678-751.
RX   MEDLINE; 79005687.
RA   FERNANDEZ H.N., HUGLI T.E.;
RL   J. BIOL. CHEM. 253:6955-6964(1978).
RN   [5]
RP   SEQUENCE OF 678-751 FROM N.A.
RX   MEDLINE; 91144547.
RA   BOHNSACK J.F., MOLLISON K.W., BUKO A.M., ASHWORTH J.C., HILL H.R.;
RL   BIOCHEM. J. 273:635-640(1991).
RN   [6]
RP   STRUCTURE BY NMR OF C5A.
RX   MEDLINE; 88309754.
RA   ZUIDERWEG E.R., MOLLISON K.W., HENKIN J., CARTER G.W.;
RL   BIOCHEMISTRY 27:3568-3580(1988).
RN   [7]
RP   STRUCTURE BY NMR OF C5A.
RX   MEDLINE; 89207527.
RA   ZUIDERWEG E.R., NETTESHEIM D.G., MOLLISON K.W., CARTER G.W.;
RL   BIOCHEMISTRY 28:172-185(1989).
RN   [8]
RP   STRUCTURE BY NMR OF C5A.
RX   MEDLINE; 89274164.
RA   ZUIDERWEG E.R., FESIK S.W.;
RL   BIOCHEMISTRY 28:2387-2391(1989).
CC   -!- FUNCTION: ACTIVATION OF C5 BY A C5 CONVERTASE INITIATES THE
CC       SPONTANEOUS ASSEMBLY OF THE LATE COMPLEMENT COMPONENTS, C5-C9,
CC       INTO THE MEMBRANE ATTACK COMPLEX. C5B HAS A TRANSIENT BINDING SITE
CC       FOR C6. THE C5B-C6 COMPLEX IS THE FOUNDATION UPON WHICH THE LYTIC
CC       COMPLEX IS ASSEMBLED.
CC   -!- SUBUNIT: C5 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 BASIC
CC       RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE
CC       BOND. C5 CONVERTASE ACTIVATES C5 BY CLEAVING THE ALPHA CHAIN,
CC       RELEASING C5A ANAPHYLATOXIN & GENERATING C5B (BETA CHAIN + ALPHA'
CC       CHAIN).
CC   -!- SIMILARITY: TO C3, C4 AND ALPHA-2-MACROGLOBULIN.
CC   -!- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C5,
CC       C5 ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT
CC       INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
CC       PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
CC       BASOPHILIC LEUKOCYTES. C5A ALSO STIMULATES THE LOCOMOTION OF
CC       POLYMORPHONUCLEAR LEUKOCYTES (CHEMOKINESIS) AND DIRECT THEIR
CC       MIGRATION TOWARD SITES OF INFLAMMATION (CHEMOTAXIS).
CC   -!- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 855
CC       ONWARD DUE TO THE PRESENCE OF AN ALU REPEAT.
CC   -!- SIMILARITY: CONTAINS ONE ANAPHYLATOXIN-LIKE DOMAIN.
DR   EMBL; M57729; G179983; -.
DR   EMBL; M65134; G179692; -.
DR   PIR; A40075; C5HU.
DR   PIR; S15121; S15121.
DR   HSSP; P01032; 1C5A.
DR   MIM; 120900; -.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN.
KW   COMPLEMENT PATHWAY; COMPLEMENT ALTERNATE PATHWAY; GLYCOPROTEIN;
KW   PLASMA; MEMBRANE ATTACK COMPLEX; CYTOLYSIS; INFLAMMATORY RESPONSE;
KW   SIGNAL; POLYMORPHISM.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19    673       COMPLEMENT C5 BETA CHAIN.
FT   PROPEP      674    677
FT   CHAIN       678   1676       COMPLEMENT C5 ALPHA CHAIN.
FT   PEPTIDE     678    751       C5A ANAPHYLATOXIN.
FT   CHAIN       752   1676       C5B (ALPHA').
FT   DOMAIN      698    732       ANAPHYLATOXIN-LIKE.
FT   DISULFID    698    724
```



```
FT   DISULFID    699    731
FT   DISULFID    711    732
FT   CARBOHYD    741    741
FT   CARBOHYD    911    911       POTENTIAL.
FT   CARBOHYD   1115   1115       POTENTIAL.
FT   CARBOHYD   1630   1630       POTENTIAL.
FT   VARIANT     518    518       F -> S.
SQ   SEQUENCE   1676 AA;  188331 MW;  9D5C6E59 CRC32;

  Query Match            100.0%;  Score 141;  DB 2;  Length 1676;
  Best Local Similarity  100.0%;  Pred. No. 1.16e-24;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Db      872 vidhqgtksskcvrqkvegss 892
            |||||||||||||||||||||
Qy        1 VIDHQGTKSSKCVRQKVEGSS 21


RESULT    2
ID   CO5_MOUSE      STANDARD;      PRT;  1680 AA.
AC   P06684;
DT   01-JAN-1988 (REL. 06, CREATED)
DT   01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   COMPLEMENT C5 PRECURSOR (CONTAINS: C5A ANAPHYLATOXIN).
GN   C5.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90153853.
RA   WETSEL R.A., FLEISCHER D.T., HAVILAND D.L.;
RL   J. BIOL. CHEM. 265:2435-2440(1990).
RN   [2]
RP   SEQUENCE OF 41-1680 FROM N.A.
RX   MEDLINE; 87185363.
RA   WETSEL R.A., OGATA R.T., TACK B.F.;
RL   BIOCHEMISTRY 26:737-743(1987).
CC   -!- FUNCTION: ACTIVATION OF C5 BY A C5 CONVERTASE INITIATES THE
CC       SPONTANEOUS ASSEMBLY OF THE LATE COMPLEMENT COMPONENTS, C5-C9,
CC       INTO THE MEMBRANE ATTACK COMPLEX. C5B HAS A TRANSIENT BINDING SITE
CC       FOR C6. THE C5B-C6 COMPLEX IS THE FOUNDATION UPON WHICH THE LYTIC
CC       COMPLEX IS ASSEMBLED.
CC   -!- SUBUNIT: C5 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 BASIC
CC       RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE
CC       BOND. C5 CONVERTASE ACTIVATES C5 BY CLEAVING THE ALPHA CHAIN,
CC       RELEASING C5A ANAPHYLATOXIN & GENERATING C5B (BETA CHAIN + ALPHA'
CC       CHAIN).
CC   -!- SIMILARITY: TO C3, C4 AND ALPHA-2-MACROGLOBULIN.
CC   -!- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C5,
CC       C5 ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT
CC       INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
CC       PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
CC       BASOPHILIC LEUKOCYTES. C5A ALSO STIMULATES THE LOCOMOTION OF
CC       POLYMORPHONUCLEAR LEUKOCYTES (CHEMOKINESIS) AND DIRECT THEIR
CC       MIGRATION TOWARD SITES OF INFLAMMATION (CHEMOTAXIS).
CC   -!- SIMILARITY: CONTAINS ONE ANAPHYLATOXIN-LIKE DOMAIN.
DR   EMBL; M35525; G309124; -.
DR   EMBL; M35526; G309123; -.
DR   PIR; A27538; A27538.
DR   PIR; A35530; A35530.
DR   HSSP; P01032; 1C5A.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN.
```

```
KW   COMPLEMENT PATHWAY; COMPLEMENT ALTERNATE PATHWAY; GLYCOPROTEIN;
KW   PLASMA; MEMBRANE ATTACK COMPLEX; CYTOLYSIS; INFLAMMATORY RESPONSE;
KW   SIGNAL.
FT   SIGNAL        1     18
FT   CHAIN        19   1680       COMPLEMENT C5.
FT   CHAIN        19    674       COMPLEMENT C5 BETA CHAIN.
FT   PROPEP      675    678
FT   CHAIN       679   1680       COMPLEMENT C5 ALPHA CHAIN.
FT   PEPTIDE     679    755       C5A ANAPHYLATOXIN.
FT   CHAIN       756   1680       C5B (ALPHA').
FT   DOMAIN      702    736       ANAPHYLATOXIN-LIKE.
FT   DISULFID    702    728       BY SIMILARITY.
FT   DISULFID    703    735       BY SIMILARITY.
FT   DISULFID    715    736       BY SIMILARITY.
FT   CARBOHYD    427    427       POTENTIAL.
FT   CARBOHYD    915    915       POTENTIAL.
FT   CARBOHYD   1119   1119       POTENTIAL.
FT   CARBOHYD   1633   1633       POTENTIAL.
FT   VARIANT     216    216       Y -> L (IN DEFECTIVE VARIANT C5D).
FT   VARIANT     217   1680       MISSING (IN DEFECTIVE VARIANT C5D).
SQ   SEQUENCE   1680 AA;  188877 MW;  AA17044B CRC32;

  Query Match            48.9%;  Score 69;  DB 2;  Length 1680;
  Best Local Similarity  47.1%;  Pred. No. 7.22e-04;
  Matches    8;  Conservative    7;  Mismatches    2;  Indels    0;  Gaps    0;

Db    880 htsrpsrcvfqriegss 896
          : :::|:|| |::||||
Qy      5 QGTKSSKCVRQKVEGSS 21


RESULT    3
ID   IROA_NEIME     STANDARD;      PRT;   943 AA.
AC   Q06379;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   IRON-REGULATED OUTER MEMBRANE PROTEIN A PRECURSOR.
GN   IROA.
OS   NEISSERIA MENINGITIDIS.
OC   PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
OC   NEISSERIACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BNVC;
RX   MEDLINE; 94011384.
RA   PETTERSSON A., VAN DER LEY P., POOLMAN J.T , TOMMASSEN J.;
RL   INFECT. IMMUN. 61:4724-4733(1993).
CC   -!- FUNCTION: UNKNOWN. MAY BE AN IRON-SIDEROPHORE RECEPTOR.
CC   -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC   -!- INDUCTION: BY IRON-STARVATION CONDITIONS.
CC   -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
DR   EMBL; X69214; G45064; -.
DR   PROSITE; PS00430; TONB_DEPENDENT_REC_1.
DR   PROSITE; PS01156; TONB_DEPENDENT_REC_2.
KW   OUTER MEMBRANE; IRON TRANSPORT; TRANSPORT; TONB BOX; SIGNAL; RECEPTOR.
FT   SIGNAL        1     27       POTENTIAL.
FT   CHAIN        28    943       IRON-REGULATED OUTER MEMBRANE PROTEIN A.
FT   SIMILAR     826    943       TONB C-TERMINAL BOX.
SQ   SEQUENCE   943 AA;  105424 MW;  16644948 CRC32;

  Query Match            41.8%;  Score 59;  DB 5;  Length 943;
  Best Local Similarity  50.0%;  Pred. No. 1.65e-01;
  Matches    7;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;
```



```
Db    595 rsrkcvprkingsn 608
          :| ||| :|::||:
Qy      8 KSSKCVRQKVEGSS 21


RESULT    4
ID   YEMA_DROME     STANDARD;      PRT;  1002 AA.
AC   P25992;
DT   01-MAY-1992 (REL. 22, CREATED)
DT   01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   YEMANUCLEIN-ALPHA.
GN   YEMA OR YG4.5.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON-S;
RX   MEDLINE; 92297435.
RA   AIT-AHMED O., BELLON B., CAPRI M., JOBLET C., THOMAS-DELAAGE M.;
RL   MECH. DEV. 37:69-80(1992).
CC   -!- FUNCTION: MAY PLAY A KEY ROLE IN EGG ORGANIZATION. IT MAY BE A
CC       TRANSCRIPTIONAL REGULATOR.
CC   -!- PTM: THE N-TERMINAL IS BLOCKED.
CC   -!- TISSUE SPECIFICITY: OOCYTE-SPECIFIC.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED AT ALL OOGENIC STAGES.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
DR   EMBL; X63503; G8838; -.
DR   PIR; S22146; S22146.
DR   FLYBASE; FBGN0005596; YEM-ALPHA.
KW   NUCLEAR PROTEIN; DNA-BINDING; REPEAT.
FT   DOMAIN       80     85       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      207    217       POLY-SER.
FT   DOMAIN      219    261       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      230    253       2 X 12 AA TANDEM REPEATS.
FT   REPEAT      230    241       1.
FT   REPEAT      242    253       2.
FT   VARIANT     698    698       S -> L (IN CDNA SEQUENCE).
SQ   SEQUENCE   1002 AA;  109310 MW;  955FD2C1 CRC32;

  Query Match            41.8%;  Score 59;  DB 11;  Length 1002;
  Best Local Similarity  50.0%;  Pred. No. 1.65e-01;
  Matches    6;  Conservative    5;  Mismatches    1;  Indels    0;  Gaps    0;

Db     47 tktakcirikld 58
          ||::||:| |::
Qy      7 TKSSKCVRQKVE 18


RESULT    5
ID   RL34_HUMAN     STANDARD;      PRT;   116 AA.
AC   P49207;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   60S RIBOSOMAL PROTEIN L34.
GN   RPL34.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=OVARY;
```

```
RX   MEDLINE; 96039267.
RA   ROMMENS J.M., DUROCHER F., MCARTHUR J., TONIN P., LEBLANC J.F.,
RA   ALLEN T., SAMSON C., FERRI L., NAROD S., MORGAN K., SIMARD J.;
RL   GENOMICS 28:530-542(1995).
CC   -!- SIMILARITY: BELONGS TO THE L34E FAMILY OF RIBOSOMAL PROTEINS.
DR   EMBL; L38941; G1008856; -.
KW   RIBOSOMAL PROTEIN.
FT   INIT_MET      0      0       BY SIMILARITY.
SQ   SEQUENCE   116 AA;  13174 MW;  490F4AF1 CRC32;

  Query Match            40.4%;  Score 57;  DB 8;  Length 116;
  Best Local Similarity  41.7%;  Pred. No. 4.57e-01;
  Matches     5;  Conservative     6;  Mismatches    1;  Indels    0;  Gaps    0;

Db     79 gsmcakcvrdri 90
          |:: :||||:::
Qy      6 GTKSSKCVRQKV 17


RESULT    6
ID   CSH_ARTSP      STANDARD;      PRT;   264 AA.
AC   P32400;
DT   01-OCT-1993 (REL. 27, CREATED)
DT   01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   N-CARBAMOYLSARCOSINE AMIDASE (EC 3.5.1.59) (N-CARBAMOYLSARCOSINE
DE   AMIDOHYDROLASE) (CSHASE).
OS   ARTHROBACTER SP.
OC   PROKARYOTA; FIRMICUTES; IRREGULAR ASPOROGENOUS RODS; CORYNEFORM GROUP.
RN   [1]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS), AND REVISIONS TO 184 AND 232.
RX   MEDLINE; 92389321.
RA   ROMAO M.J., TURK D., GOMIS-RUETH F.-X., HUBER R.;
RL   J. MOL. BIOL. 226:1111-1130(1992).
CC   -!- CATALYTIC ACTIVITY: N-CARBAMOYLSARCOSINE + H(2)O = SARCOSINE +
CC       CO(2) + NH(3).
CC   -!- SUBUNIT: HOMOTETRAMER.
CC   -!- COFACTOR: ONE SULFATE ION PER SUBUNIT.
CC   -!- PATHWAY: DEGRADATION OF CREATININE TO GLYCINE.
DR   PIR; S28969; S28969.
DR   PDB; 1NBA; 22-JUN-94.
KW   HYDROLASE; 3D-STRUCTURE.
FT   ACT_SITE    177    177       INVOLVED IN HYDROLYSIS OF THE SUBSTRATE.
SQ   SEQUENCE   264 AA;  29057 MW;  81A56865 CRC32;

  Query Match            40.4%;  Score 57;  DB 2;  Length 264;
  Best Local Similarity  40.0%;  Pred. No. 4.57e-01;
  Matches     6;  Conservative     5;  Mismatches    4;  Indels    0;  Gaps    0;

Db    171 gataagcvrhtveda 185
          |: :: |||: || :
Qy      6 GTKSSKCVRQKVEGS 20


RESULT    7
ID   SP70_DICDI     STANDARD;      PRT;   537 AA.
AC   P15269; P08126;
DT   01-AUG-1988 (REL. 08, CREATED)
DT   01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT   01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE   SPORE COAT PROTEIN SP70 PRECURSOR (BEEJIN PROTEIN).
GN   COTB.
OS   DICTYOSTELIUM DISCOIDEUM (SLIME MOLD).
OC   EUKARYOTA; PROTOZOA; SARCOMASTIGOPHORA; SARCODINA; RHIZOPODA;
```



```
OC   EUMYCETOZOA; DICTYOSTELIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90097939.
RA   FOSNAUGH K.L., LOOMIS W.F.;
RL   MOL. CELL. BIOL. 9:5215-5218(1989).
RN   [2]
RP   PRELIMINARY SEQUENCE OF 72-170 FROM N.A.
RX   MEDLINE; 87057653.
RA   GOMER R.H., DATTA S., FIRTEL R.A.;
RL   J. CELL BIOL. 103:1999-2015(1986).
DR   EMBL; M26238; G167889; -.
DR   PIR; B33485; B33485.
DR   PIR; B25439; B25439.
DR   DICTYDB; DD03009; COTB.
KW   GLYCOPROTEIN; PHOSPHORYLATION; REPEAT; SPORULATION; SIGNAL.
FT   SIGNAL        1     20
FT   CHAIN        21    537       SPORE COAT PROTEIN SP70.
FT   DOMAIN      182    250       SER/THR-RICH.
FT   DOMAIN      190    248       5.5 X 11 AA TANDEM REPEATS.
FT   REPEAT      190    200       1.
FT   REPEAT      191    211       2.
FT   REPEAT      212    222       3.
FT   REPEAT      223    233       4.
FT   REPEAT      234    244       5.
FT   REPEAT      245    248       6 (INCOMPLETE).
FT   REPEAT      256    263       PRESPORE MOTIF.
FT   REPEAT      284    291       PRESPORE MOTIF.
FT   REPEAT      364    371       PRESPORE MOTIF.
FT   CARBOHYD     97     97       POTENTIAL.
SQ   SEQUENCE   537 AA;  56650 MW;  5D59CBAC CRC32;

  Query Match            39.7%;  Score 56;  DB 9;  Length 537;
  Best Local Similarity  54.5%;  Pred. No. 7.54e-01;
  Matches     6;  Conservative     4;  Mismatches    1;  Indels    0;  Gaps    0;

Db    291 kngecirdkve 301
          |:: |:|:|||
Qy      8 KSSKCVRQKVE 18


RESULT    8
ID   GLB1_MAIZE     STANDARD;      PRT;   573 AA.
AC   P15590;
DT   01-APR-1990 (REL. 14, CREATED)
DT   01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT   01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE   GLOBULIN-1 S ALLELE PRECURSOR (GLB1-S) (7S-LIKE).
GN   GLB1.
OS   ZEA MAYS (MAIZE).
OC   EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;
OC   CYPERALES; GRAMINEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. INBRED LINE VA26;
RA   BELANGER F.C., KRIZ A.L.;
RL   PLANT PHYSIOL. 91:636-643(1989).
RN   [2]
RP   SEQUENCE OF 87-100.
RX   MEDLINE; 89374022.
RA   KRIZ A.L.;
RL   BIOCHEM. GENET. 27:239-251(1989).
CC   -!- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC       CONVICILIN, CONGLYCININ, ETC.).
```

```
CC   -!- POLYMORPHISM: THE THREE MOST COMMONLY OCCURING GLB1 ALLELES HAVE
CC       THE DESIGNATION L, I, AND S FOR LARGE, INTERMEDIATE, AND SMALL
CC       PROTEINS, RESPECTIVELY.
CC   -!- PTM: THREE PROTEIN-PROCESSING STEPS OCCUR IN THE FORMATION OF THE
CC       MATURE PROTEIN FROM THE PRIMARY TRANSLATION PRODUCT.
DR   EMBL; M24845; G168481; -.
DR   HSSP; P02853; 1CAU.
DR   MAIZEDB; 30181; -.
KW   SEED STORAGE PROTEIN; SIGNAL.
FT   SIGNAL        1     18       OR 21 (POTENTIAL).
FT   PROPEP       19     86
FT   CHAIN        87    573       GLOBULIN-1 S.
FT   CARBOHYD    349    349       POTENTIAL.
SQ   SEQUENCE   573 AA;  65029 MW;  7E755E20 CRC32;

  Query Match            39.7%;  Score 56;  DB 4;  Length 573;
  Best Local Similarity  58.3%;  Pred. No. 7.54e-01;
  Matches    7;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Db     32 hgghksgrcvrr 43
          | | ||::|||:
Qy      4 HQGTKSSKCVRQ 15


RESULT    9
ID   RPOB_PSEPU     STANDARD;      PRT;  1357 AA.
AC   P19175;
DT   01-NOV-1990 (REL. 16, CREATED)
DT   01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   DNA-DIRECTED RNA POLYMERASE BETA CHAIN (EC 2.7.7.6) (TRANSCRIPTASE
DE   BETA CHAIN) (RNA POLYMERASE BETA SUBUNIT).
GN   RPOB.
OS   PSEUDOMONAS PUTIDA.
OC   PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
OC   PSEUDOMONADACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BORODIN A.M., DANILKOVICH A.V., ALLIKMETS R.L., ROSTAPSHOV V.M.,
RA   CHERNOV I.P., AZHIKINA T.L., MONASTYRSKAYA S., SVERDLOV D.;
RL   DOKL. BIOCHEM. 302:1261-1265(1988).
RN   [2]
RP   SEQUENCE OF 1036-1357 FROM N.A.
RX   MEDLINE; 89117617.
RA   BORODIN A.M., DANILKOVICH A.V., CHERNOV I.P., AZHYKINA T.L.,
RA   ROSTAPSHOV V.M., MONASTYRSKAYA G.S.;
RL   BIOORG. KHIM. 14:1179-1182(1988).
CC   -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC       OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC       SUBSTRATES.
CC   -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC       RNA(N).
CC   -!- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
CC       ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
CC       BETA' CHAIN.
CC   -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
DR   EMBL; X15849; G45729; -.
DR   EMBL; M38319; G151547; -.
KW   TRANSCRIPTION; DNA-DIRECTED RNA POLYMERASE.
FT   CONFLICT   1180   1180       T -> N (IN REF. 2).
FT   CONFLICT   1184   1184       I -> V (IN REF. 2).
FT   CONFLICT   1236   1236       F -> S (IN REF. 2).
SQ   SEQUENCE   1357 AA;  151305 MW;  BBF88A37 CRC32;
```



```
  Query Match            39.7%;  Score 56;  DB 8;  Length 1357;
  Best Local Similarity  41.2%;  Pred. No. 7.54e-01;
  Matches    7;  Conservative    4;  Mismatches    6;  Indels    0;  Gaps    0;

Db    450 idhlgnrrvrcvgemae 466
          ||| | :  :|| :: |
Qy      2 IDHQGTKSSKCVRQKVE 18


RESULT   10
ID   YTX1_XENLA     STANDARD;      PRT;   775 AA.
AC   P14380;
DT   01-JAN-1990 (REL. 13, CREATED)
DT   01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT   01-APR-1990 (REL. 14, LAST ANNOTATION UPDATE)
DE   TRANSPOSON TX1 HYPOTHETICAL 82 KD PROTEIN (ORF 1).
OS   XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AMPHIBIA; ANURA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89384562.
RA   GARRETT J.E., KNUTZON D.S., CARROLL D.;
RL   MOL. CELL. BIOL. 9:3018-3027(1989).
DR   EMBL; M26915; G214845; -.
DR   PIR; A32494; A32494.
KW   HYPOTHETICAL PROTEIN; TRANSPOSABLE ELEMENT.
SQ   SEQUENCE   775 AA;  82355 MW;  9738B05A CRC32;

  Query Match            39.0%;  Score 55;  DB 11;  Length 775;
  Best Local Similarity  46.7%;  Pred. No. 1.24e+00;
  Matches    7;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Db    617 sntskcvssevegtp 631
          :::||||   |||::
Qy      7 TKSSKCVRQKVEGSS 21


RESULT   11
ID   COTZ_BACSU     STANDARD;      PRT;   148 AA.
AC   Q08312;
DT   01-OCT-1994 (REL. 30, CREATED)
DT   01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   SPORE COAT PROTEIN Z.
GN   COTZ.
OS   BACILLUS SUBTILIS.
OC   PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / JH642;
RX   MEDLINE; 93285989.
RA   ZHANG J., FITZ-JAMES P.C., ARONSON A.I.;
RL   J. BACTERIOL. 175:3757-3766(1993).
CC   -!- SUBCELLULAR LOCATION: SPORE OUTER COAT.
CC   -!- SUBUNIT: DISULFIDE CROSS-LINKED EITHER TO ITSELF OR TO COTY.
CC   -!- SIMILARITY: TO COTY.
DR   EMBL; L10116; G304149; -.
DR   PIR; E47119; E47119.
DR   SUBTILIST; BG10500; COTZ.
KW   SPORULATION.
SQ   SEQUENCE   148 AA;  16534 MW;  B5442F5E CRC32;

  Query Match            38.3%;  Score 54;  DB 2;  Length 148;
  Best Local Similarity  63.6%;  Pred. No. 2.01e+00;
```

```
  Matches      7;  Conservative     2;  Mismatches    2;  Indels    0;  Gaps    0;

Db       4 ktsscvreave 14
           |:| |||: ||
Qy       8 KSSKCVRQKVE 18


RESULT   12
ID   DAP2_YEAST     STANDARD;      PRT;   818 AA.
AC   P18962;
DT   01-NOV-1990 (REL. 16, CREATED)
DT   01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   DIPEPTIDYL AMINOPEPTIDASE B (EC 3.4.14.-) (DPAP B) (YSCV).
GN   DAP2 OR YHR028C.
OS   SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89174971.
RA   ROBERTS C.J., POHLIG G., ROTHMAN J.H., STEVENS T.H.;
RL   J. CELL BIOL. 108:1363-1373(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RX   MEDLINE; 94378003.
RA   JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DOVER J.,
RA   DU Z., FAVELLO A., FULTON L., GATTUNG S., GEISEL C., KIRSTEN J.,
RA   KUCABA T., HILLIER L., JIER M., JOHNSTON L., LANGSTON Y.,
RA   LATREILLE P., LOUIS E.J., MACRI C., MARDIS E., MENEZES S., MOUSER L.,
RA   NHAN M., RIFKIN L., RILES L., ST PETER H., TREVASKIS E., VAUGHAN K.,
RA   VIGNATI D., WILCOX L., WOHLDMAN P., WATERSTON R., WILSON R.,
RA   VAUDIN M.;
RL   SCIENCE 265:2077-2082(1994).
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSOSOME-LIKE
CC       VACUOLES.
CC   -!- SIMILARITY: TO DPAP A.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B; ALSO KNOWN AS THE
CC       PROLYL OLIGOPEPTIDASE FAMILY.
DR   EMBL; X15484; G3660; -.
DR   EMBL; U10399; G500698; -.
DR   PIR; A30107; A30107.
DR   PIR; S46780; S46780.
DR   LISTA; SC00265; DAP2.
DR   SGD; L0000483; DAP2.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER.
KW   HYDROLASE; DIPEPTIDASE; SERINE PROTEASE; TRANSMEMBRANE; GLYCOPROTEIN;
KW   SIGNAL-ANCHOR.
FT   DOMAIN        1     29       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     30     45       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN       46    818       LUMENAL (POTENTIAL).
FT   ACT_SITE    679    679       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    756    756       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    789    789       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD     63     63       POTENTIAL.
FT   CARBOHYD     79     79       POTENTIAL.
FT   CARBOHYD    110    110       POTENTIAL.
FT   CARBOHYD    139    139       POTENTIAL.
FT   CARBOHYD    372    372       POTENTIAL.
FT   CARBOHYD    392    392       POTENTIAL.
FT   CARBOHYD    421    421       POTENTIAL.
FT   CARBOHYD    738    738       POTENTIAL.
FT   CONFLICT     83     83       Q -> H (IN REF. 1).
FT   CONFLICT    125    125       S -> N (IN REF. 1).
```



```
FT   CONFLICT    182    188       FEEIGNE -> LRRLET (IN REF. 1).
FT   CONFLICT    200    200       D -> N (IN REF. 1).
FT   CONFLICT    366    375       TSNVVRNESS -> DFKRGKERKF (IN REF. 1).
FT   CONFLICT    808    818       AKRAFDGQFVK -> QSVLSMGNLTNELTIYSSSHRDIHKT
FT                                FSYLHTMYI (IN REF. 1).
SQ   SEQUENCE   818 AA;  93404 MW;  8D658EBB CRC32;

  Query Match             38.3%;  Score 54;  DB 3;  Length 818;
  Best Local Similarity   33.3%;  Pred. No. 2.01e+00;
  Matches      7;  Conservative     7;  Mismatches    7;  Indels    0;  Gaps    0;

Db     511 ivdfhsrkaekcdkgnvlgks 531
             ::| :: |: || : :| | |
Qy       1 VIDHQGTKSSKCVRQKVEGSS 21


RESULT   13
ID   YR71_CAEEL     STANDARD;      PRT;  1039 AA.
AC   Q09564;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 118.2 KD PROTEIN F43C1.1 IN CHROMOSOME III.
GN   F43C1.1.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   JASSAL B.;
RL   SUBMITTED (DEC-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 3.
CC   -!- SIMILARITY: CONTAINS A PP2C-LIKE DOMAIN.
DR   EMBL; Z46937; G603526; -.
DR   WORMPEP; F43C1.1; CE01582.
KW   HYPOTHETICAL PROTEIN; LEUCINE-REPEAT; REPEAT.
FT   DOMAIN      216    284       LEUCINE-RICH REPEATS.
FT   REPEAT      216    238       LRR 1.
FT   REPEAT      239    261       LRR 2.
FT   REPEAT      262    284       LRR 3.
FT   DOMAIN      669    903       PP2C-LIKE.
SQ   SEQUENCE   1039 AA;  118182 MW;  877F95CB CRC32;

  Query Match             38.3%;  Score 54;  DB 11;  Length 1039;
  Best Local Similarity   38.1%;  Pred. No. 2.01e+00;
  Matches      8;  Conservative     5;  Mismatches    7;  Indels    1;  Gaps    1;

Db     677 vsgsrgmknkqcirqvrvent 697
           |   :| |:  |:|| :|| :
Qy       1 VIDHQGTKSSKCVRQ-KVEGS 20


RESULT   14
ID   E411_ADE05     STANDARD;      PRT;   116 AA.
AC   P04489;
DT   13-AUG-1987 (REL. 05, CREATED)
DT   13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT   01-MAR-1992 (REL. 21, LAST ANNOTATION UPDATE)
DE   PROBABLE EARLY E4 11 KD PROTEIN.
OS   HUMAN ADENOVIRUS TYPE 5.
OC   VIRIDAE; DS-DNA NONENVELOPED VIRUSES; ADENOVIRIDAE; MASTADENOVIRUSES.
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RX   MEDLINE; 83164198.
RA   SARNOW P., HEARING P., ANDERSON C.W., REICH N., LEVINE A.J.;
RL   J. MOL. BIOL. 162:565-583(1982).
RN   [2]
RP   COMPLETE GENOME.
RX   MEDLINE; 92087470.
RA   CHROBOCZEK J., BIEBER F., JACROT B.;
RL   VIROLOGY 186:280-285(1992).
DR   EMBL; M73260; -; NOT_ANNOTATED_CDS.
DR   EMBL; X02998; G58502; -.
DR   PIR; B03807; Q4ADE5.
KW   EARLY PROTEIN.
SQ   SEQUENCE   116 AA;  13298 MW;  66EA9B5C CRC32;

  Query Match            37.6%;  Score 53;  DB 3;  Length 116;
  Best Local Similarity  60.0%;  Pred. No. 3.25e+00;
  Matches    6;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Db       3 rclrlkvega 12
           :|:| ||||:
Qy      11 KCVRQKVEGS 20


RESULT  15
ID   E411_ADE02     STANDARD;      PRT;   116 AA.
AC   P03241;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   01-APR-1990 (REL. 14, LAST ANNOTATION UPDATE)
DE   PROBABLE EARLY E4 11 KD PROTEIN.
OS   HUMAN ADENOVIRUS TYPE 2.
OC   VIRIDAE; DS-DNA NONENVELOPED VIRUSES; ADENOVIRIDAE; MASTADENOVIRUSES.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 82059444.
RA   HERISSE J., RIGOLET M., DUPONT DE DINECHIN S., GALIBERT F.;
RL   NUCLEIC ACIDS RES. 9:4023-4042(1981).
DR   EMBL; J01917; G209839; -.
DR   PIR; A03807; Q4ADE2.
KW   EARLY PROTEIN.
SQ   SEQUENCE   116 AA;  13255 MW;  950D6981 CRC32;

  Query Match            37.6%;  Score 53;  DB 3;  Length 116;
  Best Local Similarity  60.0%;  Pred. No. 3.25e+00;
  Matches    6;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Db       3 rclrlkvega 12
           :|:| ||||:
Qy      11 KCVRQKVEGS 20


Search completed: Wed Jan 28 12:11:02 1998
Job time : 10 secs.
```

```
******************************************************************************
MPSRCH (TM)
******************************************************************************
```

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

```
MPsrch_pp   protein - protein database search, using Smith-Waterman algorithm

Run on:          Wed Jan 28 12:09:44 1998;  MasPar time 2.31 Seconds
                                            111.852 Million cell updates/sec
Tabular output not generated.

Title:           >US-08-487-283A-1
Description:     (1-21) from US08487283A.pep
Perfect Score:   141
Sequence:        1 VIDHQGTKSSKCVRQKVEGSS 21

Scoring table:   PAM 150
                 Gap 15

Searched:        101610 seqs, 12294212 residues

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database:        a-geneseq28
                   1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
                   8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
                   14:part14 15:part15 16:part16 17:part17 18:part18
                   19:part19 20:part20 21:part21

Statistics:      Mean 18.854;  Variance 56.108;  scale 0.336
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	141	100.0	21	15	R77605	Pro-C5 polypeptide KS	1.23e-09
2	141	100.0	1676	15	R77604	Pro-C5 polypeptide.	1.23e-09
3	55	39.0	264	4	R22271	CSHase.	3.50e+01
4	54	38.3	1732	17	R96029	P. gingivalis porphyp	4.47e+01
5	54	38.3	3163	16	R94347	Hepatitis GB virus (H	4.47e+01
6	53	37.6	589	3	R14327	Mouse epithelin precu	5.69e+01
7	53	37.6	652	15	R88124	Tobacco mosaic virus	5.69e+01
8	53	37.6	1009	5	R26206	Type B human platelet	5.69e+01
9	53	37.6	1089	2	R06910	Alpha type PDGF recep	5.69e+01
10	53	37.6	1089	2	R08267	Platelet derived grow	5.69e+01
11	53	37.6	1196	19	W04326	Rat petrin.	5.69e+01
12	52	36.9	354	6	R33439	Ornithine cyclodeamin	7.25e+01
13	51	36.2	589	3	R14325	Rat epithelin precurs	9.21e+01
14	51	36.2	914	3	R15785	B.thuringiensis toxin	9.21e+01
15	51	36.2	956	3	R15784	B.thuringiensis toxin	9.21e+01
16	51	36.2	986	9	R25141	JAK2.	9.21e+01
17	51	36.2	1100	3	R15783	B.thuringiensis toxin	9.21e+01
18	51	36.2	1129	13	R70830	Murine JAK2 kinase.	9.21e+01
19	51	36.2	1144	15	R88122	Tobacco mosaic virus	9.21e+01
20	51	36.2	1144	15	R88123	Tobacco mosaic virus	9.21e+01
21	51	36.2	1588	9	R46605	Malarial PfEMP3 epito	9.21e+01
22	51	36.2	1663	9	R46608	Plasmodium falciparum	9.21e+01
23	50	35.5	1047	20	W01535	Cellular homologue of	1.17e+02
24	49	34.8	429	10	R50036	Hantaian virus Nucleo	1.48e+02
25	49	34.8	429	21	W15472	Hantaan virus nucleoc	1.48e+02
26	49	34.8	429	20	W11872	Hantavirus nuclear pr	1.48e+02
27	49	34.8	430	20	W11874	Hantavirus nuclear pr	1.48e+02
28	49	34.8	430	20	W11873	Hantavirus nuclear pr	1.48e+02
29	49	34.8	487	3	R13794	Drosophila hormone re	1.48e+02
30	49	34.8	694	1	R04107	DNA-binding protein G	1.48e+02
31	49	34.8	1323	10	R55248	N-methyl-D-aspartic a	1.48e+02
32	49	34.8	1336	12	R66041	Human N-methyl-D-aspa	1.48e+02
33	49	34.8	3080	1	P93285	Sequence of clone HIV	1.48e+02
34	48	34.0	399	18	W04218	Human urinary bladder	1.87e+02
35	48	34.0	925	14	R79148	Human insulin recepto	1.87e+02
36	48	34.0	1299	15	R86304	Drosophila patched pr	1.87e+02
37	47	33.3	122	16	R81443	Hepatitis GB virus (H	2.36e+02
38	47	33.3	458	3	R15149	CD4 coordinate system	2.36e+02
39	47	33.3	498	3	R12255	HIV-1 strain OYI GAG	2.36e+02
40	47	33.3	792	16	R85198	Avenacinase - a sapon	2.36e+02
41	47	33.3	793	16	R85200	Avenacinase-like prot	2.36e+02
42	47	33.3	793	16	R85199	Avenacinase-like prot	2.36e+02
43	47	33.3	2873	17	R87559	Hepatitis virus clone	2.36e+02
44	47	33.3	2873	17	R90796	HGV-PNF 2161 polyprot	2.36e+02
45	46	32.6	1232	19	R98217	Neuronal apoptosis in	2.97e+02

ALIGNMENTS

```
RESULT    1
ID   R77605 standard; Protein; 21 AA.
AC   R77605;
DT   02-APR-1996  (first entry)
DE   Pro-C5 polypeptide KSSKC epitope.
KW   Complement C5; haemolysis; kidney; glomerulonephritis;
KW   monoclonal antibody; antiinflammatory; antibody engineering;
KW   humanised antibody; KSSKC epitope.
OS   Homo sapiens.
PN   WO9529697-A1.
PD   09-NOV-1995.
PF   01-MAY-1995; U05688.
PR   02-MAY-1994; US-236208.
PA   (ALEX-) ALEXION PHARM INC.
PI   Evans MJ,  Matis L,  Mueller EE,  Nye SH,  Rollins S;
PI   Rother RP,  Springhorn J P,  Squinto SP,  Thomas TC;
PI   Wang Y,  Wilkins JA;
DR   WPI; 95-392923/50.
PT   Treating glomerulonephritis with antibody against complement C5
PT   component  - to inhibit complement induced cell lysis
PS   Example 13; Page 81; 181pp; English.
CC   The cDNA sequence of the complement C5 gene transcript predicts a
CC   secreted pro-C5 precursor of 1676 amino acids (R77604).  C5 is a
CC   beta-globulin heterodimer thought to play a role in the pathogenesis
CC   of glomerulonephritis (GN).  Cleavage of the C5 alpha-chain
CC   by a convertase enzyme generates anaphylatoxic C5a.  Monoclonal
```

```
CC   and humanised recombinant antibodies that recognise the alpha-chain
CC   KSSKC epitope (R77605) block C5a generation, thereby reducing
CC   glomerular inflammation and kidney dysfunction associated with GN.
SQ   Sequence   21 AA;

  Query Match            100.0%;  Score 141;  DB 15;  Length 21;
  Best Local Similarity 100.0%;  Pred. No. 1.23e-09;
  Matches    21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Db       1 vidhqgtksskcvrqkvegss 21
             |||||||||||||||||||||
Qy       1 VIDHQGTKSSKCVRQKVEGSS 21


RESULT    2
ID   R77604 standard; Protein; 1676 AA.
AC   R77604;
DT   15-MAR-1996  (first entry)
DE   Pro-C5 polypeptide.
KW   Complement C5; haemolysis; kidney; glomerulonephritis;
KW   monoclonal antibody; antiinflammatory; antibody engineering;
KW   humanised antibody.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Peptide         1..18
FT   /label= Sig_peptide
FT   Protein         19..673
FT   /label= Beta-chain
FT   Cleavage_site   673..674
FT   Cleavage_site   677..678
FT   Peptide         674..677
FT   label= Cleavage_peptide
FT   Protein         678..1676
FT   /label= Alpha-chain
FT   /note= "amino acids 872-892 (854-874 of
FT   the mature protein) comprise the KSSKS
FT   epitope"
FT   Peptide         678..751
FT   /label= C5a
FT   Cleavage_site   751..752
FT   /label= Convertase_cleavage_site
FT   Modified_site   911
FT   /label= N-glycosylation_site
FT   Modified_site   1115
FT   /label= N-glycosylation_site
FT   Modified_site   1630
FT   /label= N-glycosylation_site
PN   WO9529697-A1.
PD   09-NOV-1995.
PF   01-MAY-1995; U05688.
PR   02-MAY-1994; US-236208.
PA   (ALEX-) ALEXION PHARM INC.
PI   Evans MJ,  Matis L,  Mueller EE,  Nye SH,  Rollins S;
PI   Rother RP,  Springhorn J P,  Squinto SP,  Thomas TC;
PI   Wang Y,  Wilkins JA;
DR   WPI; 95-392923/50.
PT   Treating glomerulonephritis with antibody against complement C5
PT   component  - to inhibit complement induced cell lysis
PS   Example 13; Page 82-92; 181pp; English.
CC   The cDNA sequence of the complement C5 gene transcript predicts a
CC   secreted pro-C5 precursor of 1676 amino acids (R77604).  C5 is a
CC   beta-globulin heterodimer thought to play a role in the pathogenesis
CC   of glomerulonephritis (GN).  Cleavage of the C5 alpha-chain
CC   by a convertase enzyme generates anaphylatoxic C5a.  Monoclonal
```



```
CC   and humanised recombinant antibodies that recognise the alpha-chain
CC   KSSKC epitope (R77605) block C5a generation, thereby reducing
CC   glomerular inflammation and kidney dysfunction associated with GN.
SQ   Sequence   1676 AA;

  Query Match            100.0%;  Score 141;  DB 15;  Length 1676;
  Best Local Similarity 100.0%;  Pred. No. 1.23e-09;
  Matches    21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Db     872 vidhqgtksskcvrqkvegss 892
             |||||||||||||||||||||
Qy       1 VIDHQGTKSSKCVRQKVEGSS 21


RESULT    3
ID   R22271 standard; Protein; 264 AA.
AC   R22271;
DT   30-JUL-1992  (first entry)
DE   CSHase.
KW   N-carbamoyl-sarcosine amidohydrolase; CSH; assay; diagnosis;
KW   creatinine.
OS   Arthrobacter sp. DSM 2563.
PN   EP-476670-A.
PD   25-MAR-1992.
PF   19-SEP-1991; 115974.
PR   20-SEP-1990; DE-029844.
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
PI   Burtscher H, Schumacher G;
DR   WPI; 92-098378/13.
DR   N-PSDB; Q22713.
PT   Recombinant DNA encoding N-carbamoyl-sarcosine-amidohydrolase -
PT   useful in clinical assay of creatinine, and vectors providing
PT   efficient expression in E.coli
PS   Claim 9; Page 9 + 7; 12pp; German.
CC   The sequence encoding CSHase is useful in assay of creatinine
CC   (for diagnosis of kidney disease).  It can now be prepd. more
CC   simply than by known methods which involve culture of Arthrobacter
CC   on complex media.
SQ   Sequence   264 AA;

  Query Match             39.0%;  Score 55;  DB 4;  Length 264;
  Best Local Similarity  46.2%;  Pred. No. 3.50e+01;
  Matches     6;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Db     171 gataagcvrhtve 183
             |: :: |||: ||
Qy       6 GTKSSKCVRQKVE 18


RESULT    4
ID   R96029 standard; Protein; 1732 AA.
AC   R96029;
DT   04-SEP-1996  (first entry)
DE   P. gingivalis porphypain.
KW   Porphypain; haemagglutinin; periodontal disease; vaccine; antibody.
OS   Porphyromonas gingivalis strain W12.
FH   Key             Location/Qualifiers
FT   Region          688..708
FT   /note= "Pro-Asn repeat region type 1"
FT   Region          887..952
FT   /note= "Pro-Asn repeat region type 2"
FT   Region          946..967
FT   /note= "Pro-Asn repeat region type 1"
FT   Region          985..1006
```

```
FT   /note= "Pro-Asn repeat region type 3"
FT   Region          1041..1100
FT   /note= "Pro-Asn repeat region type 4"
FT   Region          1341..1405
FT   /note= "Pro-Asn repeat region type 2"
FT   Region          1430..1451
FT   /note= "Pro-Asn repeat region type 3"
FT   Region          1488..1547
FT   /note= "Pro-Asn repeat region type 4"
FT   Region          1607..1650
FT   /note= "Pro-Asn repeat region type 2"
PN   WO9617936-A2.
PD   13-JUN-1996.
PF   11-DEC-1995; U16108.
PR   09-DEC-1994; US-353485.
PA   (UABR-) UAB RES FOUND.
PA   (UYFL ) UNIV FLORIDA.
PI   Han N,  Lantz M,  Lepine G,  Patti JM,  Progulske-Fox A;
PI   Tumwasorn S;
DR   WPI; 96-287181/29.
DR   N-PSDB; T30653.
PT   Porphyromonas gingivalis genes and proteins - used in the detection
PT   and vaccination against periodontal disease
PS   Claim 5; Page 76-81; 153pp; English.
CC   P. gingivalis W12 cysteine protease, porphypain (R96029), was
CC   identified as the product of the prtP gene (T30653) isolated from
CC   P. gingivalis W12 genomic DNA.  The porphypain shows homology to
CC   the haemagglutinins (see also R96026-28 and R96030-33) of P.
CC   gingivalis 318.  It can be obtd. from transformed host cells and
CC   used as a vaccine to protect humans or animals against periodontal
CC   disease.  Expression in Salmonella cells allows prodn. of a live
CC   vaccine.  The porphypain and haemagglutinins can also be used to
CC   detect the presence of anti-P. gingivalis antibodies and to raise
CC   monoclonal antibodies for diagnostic appln.
SQ   Sequence   1732 AA;

  Query Match           38.3%;  Score 54;  DB 17;  Length 1732;
  Best Local Similarity  31.6%;  Pred. No. 4.47e+01;
  Matches    6;  Conservative    7;  Mismatches    6;  Indels    0;  Gaps    0;

Db   1097 itakgvrspkairgriqgt 1115
          |  | :|:| :| :::|:
Qy      2 IDHQGTKSSKCVRQKVEGS 20


RESULT   5
ID   R94347 standard; Protein; 3163 AA.
AC   R94347;
DT   02-JUL-1996  (first entry)
DE   Hepatitis GB virus (HGBV) clone GB contig A protein prod.
KW   Hepatitis GB virus; HGBV; diagnosis; treatment; vaccine;
KW   reagents; non-A; non-B; non-C; non-D; non-E; clone; GB contig A;
KW   tamarin; infected plasma; lambda phage; cDNA library.
OS   Hepatitis GB virus.
FH   Key             Location/Qualifiers
FT   Misc_difference 1..3163
FT   /note= "others correspond to degenerate or STOP
FT   codons in T00129"
PN   WO9521922-A2.
PD   17-AUG-1995.
PF   14-FEB-1995; U02118.
PR   14-FEB-1994; US-196030.
PR   13-MAY-1994; US-242654.
PR   29-JUL-1994; US-283314.
```



```
PR   23-NOV-1994; US-344190.
PR   23-NOV-1994; US-344185.
PR   27-JAN-1995; US-344557.
PA   (ABBO ) ABBOTT LAB.
PI   Buijk SL,  Dawson GJ,  Desai SM,  Erker JC,  Leary TP;
PI   Muerhoff AS,  Mushahwar IK,  Pilot-Matias TJ,  Schlauder GG;
PI   Simons JN;
DR   WPI; 95-293123/38.
DR   N-PSDB; T00129.
PT   Non-A, non-B, non-C, non-D, non-E Hepatitis virus reagents - useful
PT   for diagnosis and therapy of hepatitis GB virus
PS   Example 9; Pages 401-414; 661pp; English.
CC   Double stranded hepatitis GB virus (HGBV) DNA obtd. from HGBV
CC   infected tamarin plasma, using standard procedures, was used to
CC   prepare a lambda phage HGBV cDNA library. Clones were rescued
CC   from the lambda phage, searched against a sequence database and
CC   found to be unique HGBV sequences. The clones were then used to
CC   assemble the sequences T00129/30 (GB contig A and B) which encode
CC   the proteins R94345-47 (the 3 possible coding strand reading
CC   frames) and R82072, respectively. Reagents which comprise the HGBV
CC   DNA, or its protein prods. can be used for the diagnosis, therapy
CC   or in a vaccine to prevent HGBV infection.
SQ   Sequence   3163 AA;

  Query Match           38.3%;  Score 54;  DB 16;  Length 3163;
  Best Local Similarity  40.0%;  Pred. No. 4.47e+01;
  Matches    6;  Conservative    5;  Mismatches    4;  Indels    0;  Gaps    0;

Db    528 vvdqrplkcgscvrd 542
          |:|::  | : |||:
Qy      1 VIDHQGTKSSKCVRQ 15


RESULT   6
ID   R14327 standard; Protein; 589 AA.
AC   R14327;
DT   17-JAN-1992  (first entry)
DE   Mouse epithelin precursor.
KW   ET; growth regulation; inhibition; stimulation.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   Protein         1..589
FT   /label= precursor
FT   /note= "claim 21, page 55"
FT   Protein         280..335
FT   /label= EP-1
FT   /note= "claim 22, page 55"
FT   Protein         205..261
FT   /label= EP-2
FT   /note= "claim 23, page 55"
FT   Peptide         59..114
FT   /label= EP
FT   /note= "claim 24, page 55"
FT   Peptide         123..179
FT   /label= EP
FT   /note= "claim 25, page 55"
FT   Peptide         362..416
FT   /label= EP
FT   /note= "claim 26, page 56"
FT   Peptide         440..495
FT   /label= EP
FT   /note= "claim 27, page 56"
FT   Peptide         515..570
FT   /label= EP
```

```
FT   /note= "claim 28, page 56"
PN   WO9115510-A.
PD   17-OCT-1991.
PF   03-APR-1991; U02321.
PR   03-APR-1990; US-504508.
PR   13-MAR-1991; US-083796.
PA   (BRIM ) BRISTOL-MYERS SQUIB.
PI   Shoyab M, Plowman GD;
DR   WPI; 91-325168/44.
DR   N-PSDB; Q14340.
PT   New cysteine-rich growth modulating proteins, epithelins - useful
PT   as inhibitors of neoplastic cell growth and to promote wound
PT   healing and treat psoriasis
PS   Disclosure; Fig 23; 97pp; English.
CC   ET-1 is a bifunctional growth regulator, capable of stimulating
CC   the growth of some cell types while inhibiting the growth of others.
CC   ET-2 is functionally similar to ET-1 w.r.t. growth inhibitory
CC   bioactivity.  In contrast, however, ET-2 is apparently not capable of
CC   eliciting the growth stimulatory activity characteristic of ET-1 and,
CC   in fact, antagonises this ET-1 activity.
CC   See also Q14338-40, Q14952-53, R14328-9 and R15315-20.
SQ   Sequence   589 AA;

  Query Match            37.6%;  Score 53;  DB 3;  Length 589;
  Best Local Similarity  35.7%;  Pred. No. 5.69e+01;
  Matches    5;  Conservative    7;  Mismatches    2;  Indels    0;  Gaps    0;

Db    560 hcsargtkclrkki 573
          | :::::||:| |:
Qy      4 HQGTKSSKCVRQKV 17


RESULT   7
ID   R88124 standard; Protein; 652 AA.
AC   R88124;
DT   28-MAR-1996  (first entry)
DE   Tobacco mosaic virus resistance N gene truncated protein.
KW   Tobacco mosaic virus resistance; TMV; N gene; Solanaceae;
KW   crop improvement; transgenic plant; crop improvement.
OS   Nicotiana glutinosa.
PN   WO9535024-A1.
PD   28-DEC-1995.
PF   16-JUN-1995; U07754.
PR   17-JUN-1994; US-261663.
PA   (REGC ) UNIV CALIFORNIA.
PA   (USDA ) US SEC OF AGRIC.
PI   Baker BJ,  Whitham SA;
DR   WPI; 96-058144/06.
DR   N-PSDB; T09342.
PT   Plant virus resistance gene N sequences from tobacco - useful for
PT   generating transgenic Solanaceous plants resistant to Tobacco Mosaic
PT   Virus
PS   Claim 28; Page 75-79; 98pp; English.
CC   The Nicotiana glutinosa N gene truncated protein (R88124) mediates
CC   resistance to tobacco mosaic virus (TMV).  A cDNA clone (T09342)
CC   coding for the protein was obtd. from a N. glutinosa leaf cDNA
CC   library by transposon tagging.  DNA sequences encoding the
CC   protein can be used to generate transgenic plants, esp. Solanaceae,
CC   resistant to TMV.
SQ   Sequence   652 AA;

  Query Match            37.6%;  Score 53;  DB 15;  Length 652;
  Best Local Similarity  31.3%;  Pred. No. 5.69e+01;
  Matches    5;  Conservative    5;  Mismatches    6;  Indels    0;  Gaps    0;
```



```
Db    156 dnrdktdadcirqivd 171
          |::    : |:|| |:
Qy      3 DHQGTKSSKCVRQKVE 18


RESULT   8
ID   R26206 standard; Protein; 1009 AA.
AC   R26206;
DT   09-FEB-1993  (first entry)
DE   Type B human platelet-derived growth factor receptor.
KW   PDGF; PDGF-R; mesenchyme; tyrosine kinase; ligand binding region.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Peptide         1..23
FT   /label= Signal_peptide
FT   Protein         24..1009
FT   /label= Mature_PDGF-A
PN   WO9213867-A.
PD   20-AUG-1992.
PF   28-JAN-1992; U00730.
PR   31-JAN-1991; US-650793.
PA   (CORT-) COR THERAPEUTICS INC.
PI   Escobedo JA, Fretto LJ, Giese NA, Tomlinson JE, Williams LT;
PI   Wolf D;
DR   WPI; 92-299970/36.
DR   N-PSDB; Q27451.
PT   Platelet derived growth factor receptor (PDGF-R) poly:peptide(s)
PT   - useful as therapeutic and diagnostic agents e.g. for assaying
PT   PDGF activity in sample
PS   Disclosure; Page 90; 109pp; English.
CC   The sequence given is one allele of type A human platelet-derived
CC   growth factor (PDGF) receptor (PDGF-R).  This receptor is typically
CC   found on cells of mesenchymal origin.  It acts while in the form of
CC   two transmembrane glycoproteins, each of which is about 180 kD.
CC   This receptor has three major regions.  The first is a transmembrane
CC   region, which spans the membrane once, separating the regions of the
CC   receptor exterior to the cell from those interior to the cell.  The
CC   second region is an extracellular region which contains the domains
CC   which bind the PDGF.  The third region is an intracellular region
CC   which posseses a tyrosine kinase activity.  This tyrosine kinase
CC   domain is notable in having an insert of approx. 100 amino acids,
CC   as compared with most other receptor tyrosine kinase domains which
CC   are contiguous or have shorter insert sequences.  Fragments of this
CC   sequence between 8 and 400 amino acids comprising one or more PDGF
CC   ligand binding region from the extracellular domain may be used to
CC   bind a PDGF ligand.
SQ   Sequence   1009 AA;

  Query Match            37.6%;  Score 53;  DB 5;  Length 1009;
  Best Local Similarity  38.1%;  Pred. No. 5.69e+01;
  Matches    8;  Conservative    6;  Mismatches    7;  Indels    0;  Gaps    0;

Db    381 vddhhgstggqtvrctaegtp 401
            | ||:|: ::  ||   ||::
Qy      1 VIDHQGTKSSKCVRQKVEGSS 21


RESULT   9
ID   R06910 standard; protein; 1089 AA.
AC   R06910;
DT   16-JAN-1991  (first entry)
DE   Alpha type PDGF receptor deduced from TR4 cDNA clone.
KW   Platelet derived growth factor; T11.
```

```
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Domain          1..23
FT   /label=signal peptide
FT   Domain          24..524
FT   /label=ligand binding domain
FT   Domain          525..548
FT   /label=transmembrane region
FT   Domain          549..599
FT   /label=juxtamembrane domain
FT   Binding-site    600..627
FT   /label=ATP binding site
FT   Modified-site   849
FT   /label=tyrosine autophosphorylation site
FT   Modified-site   42..44
FT   /label=N-glycos_site
FT   Modified-site   76..78
FT   /label=N-glycos_site
FT   Modified-site   103..105
FT   /label=N-glycos_site
FT   Modified-site   179..181
FT   /label=N-glycos_site
FT   Modified-site   353..355
FT   /label=N-glycos_site
FT   Modified-site   359..361
FT   /label=N-glycos_site
FT   Modified-site   458..460
FT   /label=N-glycos_site
FT   Modified-site   468..470
FT   /label=N-glycos_site
PN   W09010013-A.
PD   07-SEP-1990.
PF   08-FEB-1990; U00617.
PR   09-FEB-1989; US-308282.
PA   (USDC ) US SEC OF COMMERCE.
PI   Matsui T, Aaronson SA, Pierce JH;
DR   WPI; 90-290306/38.
DR   N-PSDB; Q05989.
PT   Type alpha platelet-derived growth factor receptor gene - useful
PT   for transforming cells to express novel protein receptor and also
PT   susceptible to genetic engineering.
PS   Claim 7; Fig 3; 64pp; English.
CC   The TR4 clone is the largest cDNA clone related to the T11 genomic
CC   clone, isolated from a library prepd. from human thymus DNA.  The
CC   T4 cDNA clone was isolated from a M426 human embryo fibroblast
CC   cDNA library.   The coding region can be introduced into the pSV2
CC   gpt vector with a simian sarcoma virus LTR as a promoter and
CC   expressed in a host.   The resulting protein is a novel PDGF
CC   receptor designated type alpha (the known receptor is designated
CC   type beta). The polypeptide has a calculated molecular mass of 120
CC   kD and has all the characteristics of a membrane spanning tyrosine
CC   kinase receptor.  The extracellular region comprises a hydrophobic
CC   signal peptide and a ligand binding domain which has structural
CC   homology with the PDGF-R/CSF1-R subfamily.  Ten Cys residues are
CC   spaced at the same positions as in other receptors of the sub-
CC   family and eight potential N-linked glycosylation sites are also
CC   present.  A hydrophobic segment spans the membrane and the cyto-
CC   plasmic region comprises a juxtamembrane region, a tyrosine kinase
CC   region split into TK1 and TK2 by a hydrophilic interkinase region
CC   and a hydrophilic C-terminal tail.  The TK region includes the
CC   consensus ATP binding sequence (G-X-G-X-X-G...K) and a tyrosine
CC   autophosphorylation site homologous to that of pp60(v-src).
SQ   Sequence   1089 AA;
```



```
  Query Match            37.6%;  Score 53;  DB 2;  Length 1089;
  Best Local Similarity  38.1%;  Pred. No. 5.69e+01;
  Matches     8;  Conservative    6;  Mismatches    7;  Indels    0;  Gaps    0;

Db     421 vddhhgstggqtvrctaegtp 441
           | ||:|: ::  ||   ||::
Qy       1 VIDHQGTKSSKCVRQKVEGSS 21


RESULT   10
ID   R08267 standard; protein; 1089 AA.
AC   R08267;
DT   07-MAR-1991  (first entry)
DE   Platelet derived growth factor (PDGF) receptor protein.
KW   Atherosclerosis; fibrotic diseases.
OS   Homo sapiens.
PN   W09014425-A.
PD   29-NOV-1990.
PF   21-MAY-1990; U02849.
PR   22-MAY-1989; US-355018.
PA   (ZYMO-) ZYMOGENETICS INC.
PI   Kelly JD, Murray MJ;
DR   WPI; 90-375992/50.
DR   N-PSDB; Q06869.
PT   DNA encoding platelet-derived growth factor - used to transform
PT   cells for culturing to detect PDG agonists and antagonists
PS   Claim 1; Fig 1; 30pp; English.
CC   Gene product may be expressed from a transformed cell. It has
CC   utility in dection of PDGF agonist and antagonist analogues, binding
CC   AA, AB and BB isoforms. PDGF agonists may be used to enhance wound
CC   healing, and antagonists may be used to block the effects of PDGF
CC   eg. in treatment of atherosclerosis or fibrotic diseases.
SQ   Sequence   1089 AA;

  Query Match            37.6%;  Score 53;  DB 2;  Length 1089;
  Best Local Similarity  38.1%;  Pred. No. 5.69e+01;
  Matches     8;  Conservative    6;  Mismatches    7;  Indels    0;  Gaps    0;

Db     421 vddhhgstggqtvrctaegtp 441
           | ||:|: ::  ||   ||::
Qy       1 VIDHQGTKSSKCVRQKVEGSS 21


RESULT   11
ID   W04326 standard; Protein; 1196 AA.
AC   W04326;
DT   16-JAN-1997  (first entry)
DE   Rat petrin.
KW   Petrin; neurite outgrowth associated protein; CNS;
KW   central nervous system; myelin; protein phosphatase 2C; stroke;
KW   neurodegeneration.
OS   Rattus sp.
FH   Key             Location/Qualifiers
FT   Misc_difference 129
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 192
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 205
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 219
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 225
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 234
```

```
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 243
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 269
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 285
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 312
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 319
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 344
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 358
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 378
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 386
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 465
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 473
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 473
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 494
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 555
FT   /note= "corresponds to stop coodn in DNA sequence"
FT   Misc_difference 593
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 602
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 609
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 621
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 724
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 736
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 739
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 786
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 841
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 924
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 934
FT   /note= "corresponds to stop coodn in DNA sequence"
FT   Misc_difference 1017
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 1054
FT   /note= "corresponds to stop coodn in DNA sequence"
FT   Misc_difference 1127
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 1147
FT   /note= "corresponds to stop codon in DNA sequence"
FT   Misc_difference 1178
FT   /note= "corresponds to stop codon in DNA sequence"
PN   WO9632476-A1.
PD   17-OCT-1996.
PF   12-APR-1996; CA0214.
```



```
PR   13-APR-1995; US-421701.
PA   (MOUN ) MOUNT SINAI HOSPITAL CORP.
PI   Labes M,  Lozano A,  Roach A,  Roder J;
DR   WPI; 96-477127/47.
DR   N-PSDB; T38484.
PT   Assay for substance that modulates response of neuronal cells - and
PT   neurite growth associated protein, Petrin, useful in conditions
PT   involving nerve damage resulting from traumatic injury, stroke or
PT   CNS degenerative disorders
PS   Claim 9; Page 57-61; 119pp; English.
CC   Rat petrin (W04326) is a protein involved in modulating neurite
CC   growth inhibition.  The amino sequence was deduced from a cDNA
CC   clone (T38484) derived from an adult rat brain cDNA library; no
CC   coding sequence was indicated.  Petrin is a new member of the
CC   protein phosphatase 2C family, and is expressed in neurons in brain
CC   tissue, partic. in the Purkinje cells of the cerebellum.  Petrin,
CC   and antibodies raised against it, can be used to modulate neurite
CC   growth and axonal regeneration.
SQ   Sequence   1196 AA;

  Query Match            37.6%;  Score 53;  DB 19;  Length 1196;
  Best Local Similarity  61.5%;  Pred. No. 5.69e+01;
  Matches     8;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Db      900 vpshsgtksphcv 912
            |  | ||||: ||
Qy        1 VIDHQGTKSSKCV 13


RESULT   12
ID   R33439 standard; Protein; 354 AA.
AC   R33439;
DT   06-JUL-1993  (first entry)
DE   Ornithine cyclodeaminase C58 from Ti plasmid pTiC58
KW   mu-crystallins; drug targetting; nervous acting drugs; CNS; neural;
KW   neuronal; neurotransmitter agents; neuromuscular agents; NMJ;
KW   neuromuscular junctions; memory agents; Alzheimers disease;
KW   CNS depressants; CNS stimulators; tranquilisers; muscle relaxants;
KW   antispasmodics; analgesics; anesthetics; anticonvulsants;
KW   antiepileptic agents; antianxiety agents; hallucinogens; sedatives;
KW   hypnotics.
OS   Agrobacterium tumefaciens
PN   US7844304-A.
PD   01-JAN-1993.
PF   28-FEB-1992; 844304.
PR   28-FEB-1992; US-844304.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Kim R,  Wistow G;
DR   WPI; 93-093573/11.
PT   New mu-crystalline proteins - having ornithine cyclo-deaminase
PT   activity, used in diagnosis and treatment of disorders in
PT   ornithine metabolism
PS   Disclosure; Page 34; 60pp; English.
CC   This sequence represents ornithine cyclodeaminase (OCD) from
CC   Agrobacterium Ti plasmid pTiC58. It shows approximately 30%
CC   homology with the kangaroo eye lens protein mu-crystallin.
SQ   Sequence   354 AA;

  Query Match            36.9%;  Score 52;  DB 6;  Length 354;
  Best Local Similarity  63.6%;  Pred. No. 7.25e+01;
  Matches     7;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Db      312 ryvrdrvegss 322
            : ||::|||||
```

```
Qy      11 KCVRQKVEGSS 21


RESULT   13
ID   R14325 standard; Protein; 589 AA.
AC   R14325;
DT   17-JAN-1992  (first entry)
DE   Rat epithelin precursor.
KW   ET; growth regulation; inhibition; stimulation.
OS   Rattus rattus.
FH   Key             Location/Qualifiers
FT   Protein         1..589
FT   /label= precursor
FT   /note= "claim 11, page 54"
FT   Protein         280..335
FT   /label= EP-1
FT   /note= "claim 12, page 54"
FT   Protein         205..261
FT   /label= EP-2
FT   /note= "claim 13, page 54"
FT   Peptide         59..114
FT   /label= EP
FT   /note= "claim 14, page 54"
FT   Peptide         123..179
FT   /label= EP
FT   /note= "claim 15, page 54"
FT   Peptide         362..416
FT   /label= EP
FT   /note= "claim 16, page 54"
FT   Peptide         440..495
FT   /label= EP
FT   /note= "claim 17, page 54"
FT   Peptide         515..570
FT   /label= EP
FT   /note= "claim 18, page 55"
PN   WO9115510-A.
PD   17-OCT-1991.
PF   03-APR-1991; U02321.
PR   03-APR-1990; US-504508.
PR   13-MAR-1991; US-083796.
PA   (BRIM ) BRISTOL-MYERS SQUIB.
PI   Shoyab M, Plowman GD;
DR   WPI; 91-325168/44.
DR   N-PSDB; Q14338.
PT   New cysteine-rich growth modulating proteins, epithelins - useful
PT   as inhibitors of neoplastic cell growth and to promote wound
PT   healing and treat psoriasis
PS   Disclosure; Fig 18; 97pp; English.
CC   ET-1 is a bifunctional growth regulator, capable of stimulating
CC   the growth of some cell types while inhibiting the growth of others.
CC   ET-2 is functionally similar to ET-1 w.r.t. growth inhibitory
CC   bioactivity.  In contrast, however, ET-2 is apparently not capable of
CC   eliciting the growth stimulatory activity characteristic of ET-1 and,
CC   in fact, antagonises this ET-1 activity.
CC   See also Q14338-40, Q14952-53, R14328-9 and R15315-20.
SQ   Sequence   589 AA;

  Query Match             36.2%;  Score 51;  DB 3;  Length 589;
  Best Local Similarity   46.2%;  Pred. No. 9.21e+01;
  Matches    6;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Db    560 hcsakgtkclrkk 572
          | ::|::||:| |
Qy      4 HQGTKSSKCVRQK 16
```



```
RESULT   14
ID   R15785 standard; Protein; 914 AA.
AC   R15785;
DT   10-FEB-1992  (first entry)
DE   B.thuringiensis toxin/AcNPV gp64 fusion protein.
KW   chimeric; fusion protein; insecticide; AcNPV; Lepidoptera larvae;
KW   midgut targetting; bacterial endotoxin; pFAC13.
OS   Bacillus thuringiensis var. tenebriosis.
OS   Autographa californica Nuclear Polyhedrosis Virus.
PN   WO9117254-A.
PD   14-NOV-1991.
PF   02-MAY-1991; U03008.
PR   03-MAY-1990; US-518575.
PA   (REGC ) UNIV OF CALIFORNIA.
PI   Sivasubramanian N, Federici A;
DR   WPI; 91-353775/48.
DR   N-PSDB; Q14808.
PT   Extending host range or toxicity of insecticidal proteins - using
PT   protein capable of binding to gut epithelium of insects
PS   Claim 55; Fig 18; 61pp; English.
CC   A polylinker was inserted into the XmnI restriction site at the
CC   carboxyl terminus coding region of B.thuringiensis var. tenebriosis
CC   (Btt) toxin. DNA encoding the gp64 viral membrane protein of AcNPV
CC   was operably linked to the Btt toxin coding sequence via the
CC   polylinker. The gp64 gene sequences act as midgut targetting
CC   signals for bacterial endotoxins. Plasmid pFAC13 was one of three
CC   different Btt/gp64 gene fusions that were constructed and its
CC   deduced amino acid sequence is given here.
CC   See also Q14806 and Q14807.
SQ   Sequence   914 AA;

  Query Match             36.2%;  Score 51;  DB 3;  Length 914;
  Best Local Similarity   45.5%;  Pred. No. 9.21e+01;
  Matches    5;  Conservative    5;  Mismatches    1;  Indels    0;  Gaps    0;

Db    665 kfnrcikrkve 675
          | ::|:::|||
Qy      8 KSSKCVRQKVE 18


RESULT   15
ID   R15784 standard; Protein; 956 AA.
AC   R15784;
DT   10-FEB-1992  (first entry)
DE   B.thuringiensis toxin/AcNPV gp64 fusion protein.
KW   chimeric; fusion protein; insecticide; AcNPV; Lepidoptera larvae;
KW   midgut targetting; bacterial endotoxin; pFX7.
OS   Bacillus thuringiensis var. tenebriosis.
OS   Autographa californica Nuclear Polyhedrosis Virus.
PN   WO9117254-A.
PD   14-NOV-1991.
PF   02-MAY-1991; U03008.
PR   03-MAY-1990; US-518575.
PA   (REGC ) UNIV OF CALIFORNIA.
PI   Sivasubramanian N, Federici A;
DR   WPI; 91-353775/48.
DR   N-PSDB; Q14807.
PT   Extending host range or toxicity of insecticidal proteins - using
PT   protein capable of binding to gut epithelium of insects
PS   Claim 55; Fig 17; 61pp; English.
CC   A polylinker was inserted into the XmnI restriction site at the
CC   carboxyl terminus coding region of B.thuringiensis var. tenebriosis
```

```
CC   (Btt) toxin. DNA encoding the gp64 viral membrane protein of AcNPV
CC   was operably linked to the Btt toxin coding sequence via the
CC   polylinker. The gp64 gene sequences act as midgut targetting
CC   signals for bacterial endotoxins. Plasmid pFX7 was one of three
CC   different Btt/gp64 gene fusions that were constructed and its
CC   deduced amino acid sequence is given here.
CC   See also Q14806 and Q14808.
SQ   Sequence   956 AA;

  Query Match            36.2%;  Score 51;  DB 3;  Length 956;
  Best Local Similarity  45.5%;  Pred. No. 9.21e+01;
  Matches     5;  Conservative    5;  Mismatches    1;  Indels    0;  Gaps    0;

Db    707 kfnrcikrkve 717
          | ::|:::|||
Qy      8 KSSKCVRQKVE 18


Search completed: Wed Jan 28 12:09:58 1998
Job time : 14 secs.
```

```
******************************************************************************

                         MPSRCH (TM)

******************************************************************************
```

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

```
Run on:          Wed Jan 28 12:14:09 1998;  MasPar time 57.52 Seconds
                                            887.795 Million cell updates/sec
Tabular output not generated.

Title:           >US-08-487-283A-2
Description:     (1-1676) from US08487283A.pep
Perfect Score:   12048
Sequence:        1 MGLLGILCFLIFLGKTWGQE..........CQAFLANLDEFAEDIFLNGC 1676

Scoring table:   PAM 150
                 Gap 11

Searched:        95051 seqs, 30469580 residues

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database:        pir53
                   1:ann1 2:ann2 3:ann3 4:ann4 5:unann1 6:unann2 7:unann3
                   8:unann4 9:unann5 10:unann6 11:unann7 12:unann8
                   13:unann9 14:unann10 15:unann11 16:unann12 17:unenc
                   18:unrev

Statistics:      Mean 56.253;  Variance 125.273;  scale 0.449
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	12048	100.0	1676	2	C5HU	complement C5 precur	0.00e+00
2	9789	81.3	1680	2	C5MS	complement C5 precur	0.00e+00
3	2367	19.6	1651	2	C3NJ	complement C3 precur	0.00e+00
4	2361	19.6	1663	2	C3HU	complement C3 precur	0.00e+00
5	2309	19.2	1673	15	I50806	complement component	0.00e+00
6	2298	19.1	1652	6	I50711	complement C3 precur	0.00e+00
7	2296	19.1	1663	2	C3MS	complement C3 precur	0.00e+00
8	2302	19.1	1666	2	C3GP	complement C3 precur	0.00e+00
9	2255	18.7	1663	2	C3RT	complement C3 precur	0.00e+00
10	2214	18.4	1642	6	I51018	cobra venom factor p	0.00e+00
11	2106	17.5	1620	15	S21045	complement protein h	0.00e+00
12	1944	16.1	1620	6	I51339	complement component	0.00e+00
13	1692	14.0	1744	2	C4HU	complement C4A precu	5.65e-280
14	1605	13.3	1738	2	A24558	complement C4 precur	1.46e-263
15	1565	13.0	1735	6	S54784	sex-limited protein	5.00e-256
16	1544	12.8	1736	6	A29176	sex-limited protein	4.49e-252
17	796	6.6	726	6	A27602	complement C3 - rabb	3.63e-113
18	697	5.8	1476	6	JC5143	alpha-macroglobulin	2.97e-95
19	690	5.7	1487	6	S15904	alpha-1 proteinase i	5.39e-94
20	670	5.6	1477	6	A29952	alpha-1-inhibitor II	2.10e-90
21	647	5.4	1476	6	A41185	alpha-2 macroglobuli	2.77e-86
22	582	4.8	1474	2	MAHU	alpha-2-macroglobuli	1.02e-74
23	581	4.8	1482	6	S13495	pregnancy zone prote	1.53e-74
24	554	4.6	1495	6	S27001	alpha-2-macroglobuli	8.91e-70
25	525	4.4	1500	6	A42210	alpha-1-macroglobuli	1.10e-64
26	506	4.2	322	6	A32329	complement C3 - Afri	2.29e-61
27	507	4.2	1472	6	A26122	alpha-2-macroglobuli	1.54e-61
28	502	4.2	1473	6	A20872	ovostatin precursor	1.14e-60
29	493	4.1	1451	6	B41185	alpha-2 macroglobuli	4.20e-59
30	470	3.9	1464	6	JC5144	murinoglobulin - gui	4.05e-55
31	456	3.8	785	6	A29953	alpha-1-inhibitor II	1.05e-52
32	399	3.3	74	6	A01268	complement C5 - pig	5.47e-43
33	385	3.2	74	6	A25408	complement C5 - bovi	1.24e-40
34	370	3.1	77	6	A57689	complement C5a - rat	4.02e-38
35	349	2.9	594	6	I48771	Slp(w7) - mouse (fra	1.23e-34
36	354	2.9	643	6	I39375	alpha-2-macroglobuli	1.83e-35
37	319	2.6	609	6	A41081	alpha-1-inhibitor II	9.98e-30
38	267	2.2	170	6	I84442	complement C4 - oran	1.92e-21
39	257	2.1	170	6	I61887	complement C4 - oran	6.84e-20
40	255	2.1	170	6	I61888	complement C4 - chim	1.39e-19
41	251	2.1	170	6	I37024	complement C4 - gori	5.74e-19
42	249	2.1	170	6	I61890	complement C4 - chim	1.16e-18
43	237	2.0	163	6	S33348	complement C4 protei	7.78e-17
44	235	2.0	163	6	S33345	complement C4 protei	1.56e-16
45	217	1.8	163	6	S33344	complement C4 protei	7.49e-14

ALIGNMENTS

```
RESULT     1
ENTRY           C5HU      #type complete
TITLE           complement C5 precursor - human
CONTAINS        C5a anaphylatoxin; C5b
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change
                  27-Feb-1997
ACCESSIONS      A40075; A27689; A01267; A01266; S15121
REFERENCE       A40075
   #authors     Haviland, D.L.; Haviland, J.C.; Fleischer, D.T.; Hunt, A.;
                  Wetsel, R.A.
   #journal     J. Immunol. (1991) 146:362-368
   #title       Complete cDNA sequence of human complement pro-C5. Evidence
                  of truncated transcripts derived from a single copy gene.
   #cross-references MUID:91079575
   #accession   A40075
      ##molecule_type mRNA
      ##residues     1-1676 ##label HAV
      ##cross-references GB:M57729
      ##note         518-Ser was also found
REFERENCE       A27689
   #authors     Wetsel, R.A.; Lemons, R.S.; Le Beau, M.M.; Barnum, S.R.;
                  Noack, D.; Tack, B.F.
   #journal     Biochemistry (1988) 27:1474-1482
   #title       Molecular analysis of human complement component C5:
                  localization of the structural gene to chromosome 9.
```

```
   #cross-references MUID:88209511
   #accession    A27689
      ##molecule_type mRNA
      ##residues     412-1676 ##label WET
      ##cross-references GB:M18879
REFERENCE        A01267
   #authors      Fernandez, H.N.; Hugli, T.E.
   #journal      J. Biol. Chem. (1978) 253:6955-6964
   #title        Primary structural analysis of the polypeptide portion of
                   human C5a anaphylatoxin. Polypeptide sequence determination
                   and assignment of the oligosaccharide attachment site in
                   C5a.
   #cross-references MUID:79005687
   #accession    A01267
      ##molecule_type protein
      ##residues     678-751 ##label FER
REFERENCE        A01266
   #authors      Lundwall, A.B.; Wetsel, R.A.; Kristensen, T.; Whitehead,
                   A.S.; Woods, D.E.; Ogden, R.C.; Colten, H.R.; Tack, B.F.
   #journal      J. Biol. Chem. (1985) 260:2108-2112
   #title        Isolation and sequence analysis of a cDNA clone encoding the
                   fifth complement component.
   #cross-references MUID:85130937
   #accession    A01266
      ##molecule_type mRNA
      ##residues     412-854,
                      'SLALSPRLECNGKISGHCKLRLPGSSDSPASASQVAGITGTHHHAQPT'
                      ##label LUN
      ##cross-references GB:K02874
      ##note         the carboxyl-terminal part of the sequence in this
                       report appears to be derived from translation of an
                       ALU repeat sequence
REFERENCE        S15121
   #authors      Bohnsack, J.F.; Mollison, K.W.; Buko, A.M.; Ashworth, J.C.;
                   Hill, H.R.
   #journal      Biochem. J. (1991) 273:635-640
   #title        Group B streptococci inactivate complement component C5a by
                   enzymic cleavage at the C-terminus.
   #cross-references MUID:91144547
   #contents     annotation
COMMENT    Complement C5 contains two disulfide-linked chains, formed by
             removal of four basic residues. C5 convertase releases C5a
             anaphylatoxin from the amino end of the alpha chain, generating
             C5b (beta and alpha' chains).
COMMENT    Activation of C5 initiates the spontaneous assembly of the late
             complement components, C5-C9, into the membrane attack complex.
             C5b has a transient binding site for C6. The C5b-C6 complex is
             the foundation upon which the membrane attack complex is
             assembled.
COMMENT    C5a has potent spasmogenic and chemotactic activity.
GENETICS
   #gene         GDB:C5
      ##cross-references GDB:119734
   #map_position 9q33-9q33
CLASSIFICATION   #superfamily alpha-2-macroglobulin
KEYWORDS         complement alternate pathway; complement pathway; cytolysis;
                   glycoprotein; inflammation; membrane attack complex; plasma
FEATURE
   1-18               #domain signal sequence #status predicted #label SIG\
   19-673,678-1676    #product complement C5 #status predicted #label MAT\
   19-673,752-1676    #product C5b #status predicted #label C5B\
   19-673             #product complement C5 and C5b beta chain #status
                        predicted #label C5BB\
   678-1676           #product complement C5 alpha chain #status predicted
```



```
                      #label C5A\
   678-751            #product C5a anaphylatoxin #status experimental #label
                        C5T\
   752-1676           #product C5b alpha' chain #status predicted #label C5BA\
   567-810,634-669,
   698-724,699-731,
   711-732,866-1527,
   1101-1159,
   1375-1505,
   1405-1474,
   1520-1525,
   1532-1606,
   1553-1676,
   1654-1657          #disulfide_bonds #status predicted\
   741                #binding_site carbohydrate (Asn) (covalent) #status
                        experimental\
   751-752            #cleavage_site Arg-Leu (C5 convertase) #status
                        experimental\
   911,1115,1630      #binding_site carbohydrate (Asn) (covalent) #status
                        predicted
SUMMARY          #length 1676  #molecular-weight 188330  #checksum 3858

  Query Match            100.0%;  Score 12048;  DB 2;  Length 1676;
  Best Local Similarity  100.0%;  Pred. No. 0.00e+00;
  Matches  1676;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Db          1 mgllgilcflifll gktwgqeqtyvisapkifrvgaseniviqvygyteafdatisiksyp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy          1 MGLLGILCFLIFLGKTWGQEQTYVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSYP 60

Db         61 dkkfsyssghvhlssenkfqnsailtiqpkqlpggqnpvsyvylevvskhfskskrmpit 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy         61 DKKFSYSSGHVHLSSENKFQNSAILTIQPKQLPGGQNPVSYVYLEVVSKHFSKSKRMPIT 120

Db        121 ydngflfihtdkpvytpdqsvkvrvyslnddlkpakretvltfidpegsevdmveeidhi 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy        121 YDNGFLFIHTDKPVYTPDQSVKVRVYSLNDDLKPAKRETVLTFIDPEGSEVDMVEEIDHI 180

Db        181 giisfpdfkipsnprygmwtikakykedfsttgtayfevkeyvlphfsvsiepeynfigy 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy        181 GIISFPDFKIPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEPEYNFIGY 240

Db        241 knfknfeitikaryfynkvvteadvyitfgiredlkddqkemmqtamqntmlingiaqvt 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy        241 KNFKNFEITIKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLINGIAQVT 300

Db        301 fdsetavkelsyysledlnnkylyiavtviestggfseeaeipgikyvlspyklnlvatp 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy        301 FDSETAVKELSYYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLNLVATP 360

Db        361 lflkpgipypikvqvkdsldqlvggvpvilnaqtidvnqetsdldpsksvtrvddgvasf 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy        361 LFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDDGVASF 420

Db        421 vlnlpsgvtvlefnvktdapdlpeenqaregyraiayssl sqsylyidwtdnhkallvge 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy        421 VLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKALLVGE 480

Db        481 hlniivtpkspyidkithynylilskgkiihfgtrekfsdasyqsinipvtqnmvpssrl 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy        481 HLNIIVTPKSPYIDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQNMVPSSRL 540

Db        541 lvyyivtgeqtaelvsdsvwlnieekcgnqlqvhlspdadayspgqtvslnmatgmdswv 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy     541 LVYYIVTGEQTAELVSDSVWLNIEEKCGNQLQVHLSPDADAYSPGQTVSLNMATGMDSWV 600

Db     601 alaavdsavygvqrgakkplervfqfleksdlgcgagggInnanvfhlagltfltnanad 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     601 ALAAVDSAVYGVQRGAKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLAGLTFLTNANAD 660

Db     661 dsqendepckeilrprrtlqkkieeiaakykhsvvkkccydgacvnndetceqraarisl 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     661 DSQENDEPCKEILRPRRTLQKKIEEIAAKYKHSVVKKCCYDGACVNNDETCEQRAARISL 720

Db     721 gprcikafteccvvasqlranishkdmqlgrlhmktllpvskpeirsyfpeswlwevhlv 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     721 GPRCIKAFTECCVVASQLRANISHKDMQLGRLHMKTLLPVSKPEIRSYFPESWLWEVHLV 780

Db     781 prrkqlqfalpdslttweiqgigisntgicvadtvkakvfkdvflemnipysvvrgeqiq 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     781 PRRKQLQFALPDSLTTWEIQGIGISNTGICVADTVKAKVFKDVFLEMNIPYSVVRGEQIQ 840

Db     841 lkgtvynyrtsgmqfcvkmsavegictsespvidhqgtkssKcvrqkvegssshlvtftv 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     841 LKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTKSSKCVRQKVEGSSSHLVTFTV 900

Db     901 lpleiglhninfsletwfgkeilvktlrvvpegvkresysgvtldprgiygtisrrkefp 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     901 LPLEIGLHNINFSLETWFGKEILVKTLRVVPEGVKRESYSGVTLDPRGIYGTISRRKEFP 960

Db     961 yripldlvpkteikrilsvkgllvgeilsavlsqeginilthlpkgsaeaelmsvvpvfy 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     961 YRIPLDLVPKTEIKRILSVKGLLVGEILSAVLSQEGINILTHLPKGSAEAELMSVVPVFY 1020

Db    1021 vfhyletgnhwnifhsdpliekqklkkklkegmlsimsyrnadysysvwkggsastwlta 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1021 VFHYLETGNHWNIFHSDPLIEKQKLKKKLKEGMLSIMSYRNADYSYSVWKGGSASTWLTA 1080

Db    1081 falrvlgqvnkyveqnqnsicnsllwlvenyqldngsfkensqyqpiklqgtlpvearen 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1081 FALRVLGQVNKYVEQNQNSICNSLLWLVENYQLDNGSFKENSQYQPIKLQGTLPVEAREN 1140

Db    1141 slyltaftvigirkafdicplvkidtalikadnfllentlpaqstftlaisayalslgdk 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1141 SLYLTAFTVIGIRKAFDICPLVKIDTALIKADNFLLENTLPAQSTFTLAISAYALSLGDK 1200

Db    1201 thpqfrsivsalkrealvkgnppiyrfwkdnlqhkdssvpntgtarmvettayalltsln 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1201 THPQFRSIVSALKREALVKGNPPIYRFWKDNLQHKDSSVPNTGTARMVETTAYALLTSLN 1260

Db    1261 lkdinyvnpvikwlseeqryggrfystqdtinaiegIteysllvkqlrlsmdidvsykhk 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1261 LKDINYVNPVIKWLSEEQRYGGGFYSTQDTINAIEGLTEYSLLVKQLRLSMDIDVSYKHK 1320

Db    1321 galhnykmtdknflgrpvevllnddlivstgfgsglatvhvttvvhktstseevcsfylk 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1321 GALHNYKMTDKNFLGRPVEVLLNDDLIVSTGFGSGLATVHVTTVVHKTSTSEEVCSFYLK 1380

Db    1381 idtqdieashyrgygnsdykrivacasykpsreesssgsshavmdislptgisaneedlk 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1381 IDTQDIEASHYRGYGNSDYKRIVACASYKPSREESSSGSSHAVMDISLPTGISANEEDLK 1440

Db    1441 alvegvdqlftdyqikdghvilqlnsipssdflcvrfrifelfevgflspatftvyeyhr 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1441 ALVEGVDQLFTDYQIKDGHVILQLNSIPSSDFLCVRFRIFELFEVGFLSPATFTVYEYHR 1500

Db    1501 pdkqctmfystsnikiqkvcegaackcveadcgqmqeeldltisaetrkqtackpeiaya 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
Qy    1501 PDKQCTMFYSTSNIKIQKVCEGAACKCVEADCGQMQEELDLTISAETRKQTACKPEIAYA 1560

Db    1561 ykvsitsitvenvfvkykatlldiyktgeavaekdseitfikkvtctnaelvkgrqylim 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1561 YKVSITSITVENVFVKYKATLLDIYKTGEAVAEKDSEITFIKKVTCTNAELVKGRQYLIM 1620

Db    1621 gkealqikynfsfryiypldsltwieywprdttcsscqaflanldefaediflngc 1676
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1621 GKEALQIKYNFSFRYIYPLDSLTWIEYWPRDTTCSSCQAFLANLDEFAEDIFLNGC 1676


RESULT      2
ENTRY             C5MS      #type complete
TITLE             complement C5 precursor - mouse
CONTAINS          C5a anaphylatoxin; C5b
ORGANISM          #formal_name Mus musculus #common_name house mouse
DATE              19-Nov-1988 #sequence_revision 15-Oct-1994 #text_change
                    16-Feb-1997
ACCESSIONS        A35530; A27538; A40429
REFERENCE         A35530
   #authors       Wetsel, R.A.; Fleischer, D.T.; Haviland, D.L.
   #journal       J. Biol. Chem. (1990) 265:2435-2440
   #title         Deficiency of the murine fifth complement component (C5). A
                    2-base pair gene deletion in a 5'-exon.
   #cross-references MUID:90153853
   #accession     A35530
      ##molecule_type mRNA
      ##residues      1-215,'L' ##label WET
      ##cross-references GB:J05234
REFERENCE         A27538
   #authors       Wetsel, R.A.; Ogata, R.T.; Tack, B.F.
   #journal       Biochemistry (1987) 26:737-743
   #title         Primary structure of the fifth component of murine
                    complement.
   #cross-references MUID:87185363
   #accession     A27538
      ##molecule_type mRNA
      ##residues      'PGL',44-1680 ##label WET2
REFERENCE         A40429
   #authors       Haviland, D.L.; Haviland, J.C.; Fleischer, D.T.; Wetsel, R.A.
   #journal       J. Biol. Chem. (1991) 266:11818-11825
   #title         Structure of the murine fifth complement component (C5) gene.
                    A large, highly interrupted gene with a variant donor
                    splice site and organizational homology with the third and
                    fourth complement component genes.
   #cross-references MUID:91268053
   #accession     A40429
      ##molecule_type DNA
      ##residues      1-15 ##label HAV
      ##cross-references GB:M64852
COMMENT     Complement C5 contains two disulfide-linked chains, formed by
              removal of four basic residues. C5 convertase releases C5a
              anaphylatoxin from the amino end of the alpha chain, generating
              C5b (beta and alpha' chains).
COMMENT     Activation of C5 initiates the spontaneous assembly of the late
              complement components, C5-C9, into the membrane attack complex.
              C5b has a transient binding site for C6. The C5b-C6 complex is
              the foundation upon which the membrane attack complex is
              assembled.
COMMENT     C5a has potent spasmogenic and chemotactic activity.
GENETICS
   #map_position 2
   #introns       22/3; 86/3; 140/3; 164/3; 195/2; 223/1; 253/2; 291/3; 334/1;
                    372/3; 434/3; 502/3; 572/3; 622/3; 667/1; 691/1; 757/1;
```

```
                     787/2; 812/1; 858/3; 934/3; 955/1; 985/1; 1056/1; 1081/2;
                     1134/3; 1166/3; 1224/1; 1292/3; 1343/3; 1364/3; 1392/1;
                     1411/2; 1445/3; 1470/3; 1506/1; 1534/1; 1564/1; 1592/1;
                     1637/2
CLASSIFICATION     #superfamily alpha-2-macroglobulin
KEYWORDS           complement alternate pathway; complement pathway; cytolysis;
                     glycoprotein; inflammation; membrane attack complex; plasma
FEATURE
   1-18                #domain signal sequence #status predicted #label SIG\
   19-674,679-1679     #product complement C5 #status predicted #label MAT\
   19-674,756-1679     #product C5b #status predicted #label C5B\
   19-674              #product complement C5 and C5b beta chain #status
                         predicted #label C5BB\
   679-1679            #product complement C5 alpha chain #status predicted
                         #label C5A\
   679-755             #product C5a anaphylatoxin #status predicted #label C5T\
   756-1679            #product C5b alpha' chain #status predicted #label C5BA\
   567-814,635-670,
   702-728,703-735,
   715-736,870-1531,
   1105-1163,
   1379-1509,
   1409-1478,
   1524-1529,
   1536-1609,
   1557-1679,
   1657-1660           #disulfide_bonds #status predicted\
   915,1119,1633       #binding_site carbohydrate (Asn) (covalent) #status
                         predicted
SUMMARY            #length 1680  #molecular-weight 188876  #checksum 3888

  Query Match          81.3%;  Score 9789;  DB 2;  Length 1680;
  Best Local Similarity  77.8%;  Pred. No. 0.00e+00;
  Matches 1307;  Conservative  218;  Mismatches 149;  Indels   7;  Gaps   6;

Db       1 mglwgilcllifldktwgqeqtyvisapkilrvgssenvviqvhgyteafdatlslksyp 60
           ||| ||||:|||| |||||||||||||||:|||:|||:|||| |||||||||:|:||||
Qy       1 MGLLGILCFLIFLGKTWGQEQTYVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSYP 60

Db      61 dkkvtfssgyvnlspenkfqnaalltlqpnqvpreespvshvylevvskhfskskkipit 120
           ||| ::||| |:||:||||||:|:|||:|:|:|  ::||| |||||||||||||:::|||
Qy      61 DKKFSYSSGHVHLSSENKFQNSAILTIQPKQLPGGQNPVSYVYLEVVSKHFSKSKRMPIT 120

Db     121 ynngilfihtdkpvytpdqsvkirvyslgddlkpakretvltfidpegsevdiveendyt 180
           |:||:|||||||||||||||||:|||||:||||| ||||||||||||||||||:||| |
Qy     121 YDNGFLFIHTDKPVYTPDQSVKVRVYSLNDDLKPAKRETVLTFIDPEGSEVDMVEEIDHI 180

Db     181 giisfpdfkipsnpkygvwtikanykkdfttgtayfeikeyvlprfsvsielertfigy 240
           |||||||||||||:||:|||||:|| ||:|||||||||:||||||:|||||| |  ||||
Qy     181 GIISFPDFKIPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEPEYNFIGY 240

Db     241 knfknfeitvkaryfynkvvpdaevyaffglredikdeekqmmhkatqaaklvdgvaqis 300
           ||||||||:|||||||||| :|:||  ||:|||:|::|:||: | | ::|::|:|||::
Qy     241 KNFKNFEITIKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLINGIAQVT 300

Db     301 fdsetavkelsynsledlnnkylyiavtvtessggfseeaeipgvkyvlspytlnlvatp 360
           ||||||||||| ||||||||||||||||| ||:||||||||||:|||||||| |||||||
Qy     301 FDSETAVKELSYYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLNLVATP 360

Db     361 lfvkpgipfsikaqvkdsleqavggvpvtlmaqtvdvnqetsdletkrsithdtdgvavf 420
           ||:|||||::|| ||||||:| ||||| | |||:||||||||:  :|:|:  |||| |
Qy     361 LFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDDGVASF 420

Db     421 vlnlpsnvtvlkfeirtddpelpeenqaskeyeavayssl sqsyiyiawtenykpmlvge 480
```



```
           |||||| |||| |:::|| |:|||||||   | |:||||||||||:|| ||:| |::||||
Qy     421 VLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKALLVGE 480

Db     481 ylnimvtpkspyidkithynylilskgkivqygtreklfsstyqninipvtqnmvpsarl 540
            |||:||||||||||||||||||||||||||:::|||||:  ::||:||||||||||||:||
Qy     481 HLNIIVTPKSPYIDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQNMVPSSRL 540

Db     541 lvyyivtgeqtaelvadavwinieekcgnqlqvhlspdeyvyspgqtvsldmvteadswv 600
           ||||||||||||||:|:||:|||||||||||||||||||||    |||||||||:| |  ||||
Qy     541 LVYYIVTGEQTAELVSDSVWLNIEEKCGNQLQVHLSPDADAYSPGQTVSLNMATGMDSWV 600

Db     601 alsavdravykvqgnakramqrvfqaldeksdlgcgagghdnadvfhlagltfltnana 660
           ||:||| ||| || ||::::||||| | |||||||||||| :||:|||||||||||||||
Qy     601 ALAAVDSAVYGVQRGAKKPLERVFQFL-EKSDLGCGAGGGLNNANVFHLAGLTFLTNANA 659

Db     661 ddshyrddsckeilrskrnlhllrqkieeqaakykhsvpkkccydgarvnfyetceerva 720
           |||:  |::||||||::| |:  : |||| |||||||| |||||||| ||  ||||:| |
Qy     660 DDSQENDEPCKEILRPRRTLQ--K-KIEEIAAKYKHSVVKKCCYDGACVNNDETCEQRAA 716

Db     721 rvtigplcirafnecctiankirkesphkpvqlgrihiktllpvmkadirsyfpeswlwe 780
           |:::|| ||:|| ||| :|: :| : :|| :||||:|:|||||| |::|||||||||||||
Qy     717 RISLGPRCIKAFTECCVVASQLRANISHKDMQLGRLHMKTLLPVSKPEIRSYFPESWLWE 776

Db     781 ihrvpkrkqlqvtlpdslttweiqgigisdngicvadtlkakvfkevflemnipysvvrg 840
           :| ||:||||| :||||||||||||||: ||||||:||||||:|||||||||||||||
Qy     777 VHLVPRRKQLQFALPDSLTTWEIQGIGISNTGICVADTVKAKVFKDVFLEMNIPYSVVRG 836

Db     841 eqiqlkgtvynymtsgtkfcvkmsavegictsgssaaslhtsrpsrcvfqriegssshlv 900
           |||||||||||| |||  ||||||||||||| |:     : :::|:|| |::||||||||
Qy     837 EQIQLKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTKSSKCVRQKVEGSSSHLV 896

Db     901 tftllpleiglhsinfsletsfgkdilvktlrvvpegvkresyagvildpkgirgivnrr 960
           |||:|||||||:||||||| |||:|||||||||||||||||||:|| |||:|| | ::||
Qy     897 TFTVLPLEIGLHNINFSLETWFGKEILVKTLRVVPEGVKRESYSGVTLDPRGIYGTISRR 956

Db     961 kefpyripldlvpktkverilsvkgllvgeflstvlskeginilthlpkgsaeaelmsia 1020
           |||||||||||||||||  :  ||||||||||||:|||:||| ||||||||||||||||||:
Qy     957 KEFPYRIPLDLVPKTEIKRILSVKGLLVGEILSAVLSQEGINILTHLPKGSAEAELMSVV 1016

Db    1021 pvfyvfhyleagnhwnifypdtlskrqslekkikqgvvsvmsyrnadysysmwkgasast 1080
           ||||||||||||| :| |   :| | ||:|::|:||:||||||||||||:||||:||||
Qy    1017 PVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEGMLSIMSYRNADYSYSVWKGGSAST 1076

Db    1081 wltafalrvlgqvakyvkqgdensicnsllwlvekcqlengsfkensqylpiklqgtlpae 1140
           |||||||||||| |||| |::||||||||||||||: ||:||||||||| ||||||||| |
Qy    1077 WLTAFALRVLGQVNKYVEQNQNSICNSLLWLVENYQLDNGSFKENSQYQPIKLQGTLPVE 1136

Db    1141 aqektlyltafsvigirkavdicptmkihtaldkadsfllentlpskstftlaivayals 1200
           |:|::||||||:|||||||| ||| :|| ||| |||:||||||||: |||||| |||||
Qy    1137 ARENSLYLTAFTVIGIRKAFDICPLVKIDTALIKADNFLLENTLPAQSTFTLAISAYALS 1196

Db    1201 lgdrthprfrlivsalrkeafvkgdppiyrywrdtlkrpdssvpssgtagmvettayall 1260
           |||:|||:|| |||||::|:|||:|||||:|:| |  : |||||::||| |||||||||
Qy    1197 LGDKTHPQFRSIVSALKREALVKGNPPIYRFWKDNLQHKDSSVPNTGTARMVETTAYALL 1256

Db    1261 aslklkdmnyanpiikwlseeqryggg fystqdtinaieglteyslllkqihldmdinva 1320
           :||:|||:|| ||:|||||||||||||||||||||||||||||||||||||:||::| |||:|:
Qy    1257 TSLNLKDINYVNPVIKWLSEEQRYGGGFYSTQDTINAIEGLTEYSLLVKQLRLSMDIDVS 1316

Db    1321 ykhegdfhkykvtekhflgrpvevslnddlvvstgyssglatvyvktvvhkisvseefcs 1380
           ||| |  :|:||:|:|:||||||| |||||:||||::|||||| | ||||| | ||| ||
Qy    1317 YKHKGALHNYKMTDKNFLGRPVEVLLNDDLIVSTGFGSGLATVHVTTVVHKTSTSEEVCS 1376

Db    1381 fylkidtqdieasshfr-lsdsgfkriiacasykpskeestsgsshavmdislptgigan 1439
```

```
            ||||||||||| |:|  ::| :|||:||||||||:|||:|||||||||||||||:||
Qy   1377 FYLKIDTQDIEAS-HYRGYGNSDYKRIVACASYKPSREESSSGSSHAVMDISLPTGISAN 1435

Db   1440 eedlralvegvdqlltdyqikdghvilqlnsipsrdflcvrfrifelfqvgflnpatftv 1499
          ||||:||||||||:|||||||||||||||||||| |||||||||||||:||||:||||||
Qy   1436 EEDLKALVEGVDQLFTDYQIKDGHVILQLNSIPSSDFLCVRFRIFELFEVGFLSPATFTV 1495

Db   1500 yeyhrpdkqctmiysisdtrlqkvcegaactcveadcaqlqaevdlaisadsrkekackp 1559
          |||||||||||:|| |: ::|||||||| |||||:|:| |:||:|||::||: ||||
Qy   1496 YEYHRPDKQCTMFYSTSNIKIQKVCEGAACKCVEADCGQMQEELDLTISAETRKQTACKP 1555

Db   1560 etayaykvritsateenvfvkytatllvtyktgea-adensevtfikkmsctnanlvkgk 1618
          | |||||| ||| | ||||||| ||||  |||||| |: :||:|||||::||||:||||:
Qy   1556 EIAYAYKVSITSITVENVFVKYKATLLDIYKTGEAVAEKDSEITFIKKVTCTNAELVKGR 1615

Db   1619 qylimgkevlqikhnfsfkyiyplddsstwieywptdttcpscqafvenlnnfaedlflns 1678
          ||||||| |||| ||||:||||||| |||||||| ||||:|||||: ||::||||:|||:
Qy   1616 QYLIMGKEALQIKYNFSFRYIYPLDSLTWIEYWPRDTTCSSCQAFLANLDEFAEDIFLNG 1675

Db   1679 c 1679
          |
Qy   1676 C 1676


RESULT    3
ENTRY          C3NJ        #type complete
TITLE          complement C3 precursor - Naja naja
CONTAINS       alternative-complement-pathway C3/C5 convertase (EC
                 3.4.21.47) C3b subunit; C3a anaphylatoxin
ORGANISM       #formal_name Naja naja
DATE           18-Jun-1993 #sequence_revision 07-Oct-1994 #text_change
                 16-Feb-1997
ACCESSIONS     A46513
REFERENCE      A46513
   #authors    Fritzinger, D.C.; Petrella, E.C.; Connelly, M.B.; Bredehorst,
                 R.; Vogel, C.W.
   #journal    J. Immunol. (1992) 149:3554-3562
   #title      Primary structure of cobra complement component C3.
   #cross-references MUID:93056528
   #accession     A46513
      ##molecule_type mRNA
      ##residues       1-1651 ##label FRI
      ##cross-references NCBIP:118403; GB:L02365
      ##note           authors' translation shows Arg-1408 after residue 1438
                         and, consequently, residues 1408-1438 are displaced
                         one codon to the left
      ##note           sequence extracted from NCBI backbone and corrected to
                         correspond with the translation of the nucleotide
                         sequence
COMMENT     Complement C3 contains two chains, formed by removal of four
              residues and linked by a disulfide bond. Its activation by a C3
              convertase, which is the central reaction in both classical and
              alternative complement pathways, releases the C3a anaphylatoxin
              from the amino end of the alpha chain and generates C3b, which
              associates with the Bb fragment of complement factor B to form
              the alternative-complement-pathway C3/C5 convertase.
COMMENT     C3a anaphylatoxin is a vasoactive peptide and a mediator of
              inflammation.
COMMENT     C3b, with its highly reactive thiol group, binds to the surface of
              foreign particles and facilitates phagocytosis. It binds to
              complement C5 and renders it susceptible to proteolysis by the
              classical-complement-pathway C3/C5 convertase. The activity of
              C3b is regulated by proteolytic cleavage involving factors H and
              I. Its degradation products can also be biologically active.
```



```
COMMENT    The major site of synthesis of this plasma protein is the liver.
CLASSIFICATION  #superfamily alpha-2-macroglobulin
KEYWORDS        acute phase; complement alternate pathway; complement
                  pathway; glycoprotein; hydrolase; immune response;
                  inflammation; liver; plasma; serine proteinase; thiolester
                  bond
FEATURE
   1-22               #domain signal sequence #status predicted #label SIG\
   23-655             #product complement C3 and C3b beta chain #status
                        predicted #label C3BB\
   23-655,660-1651    #product complement C3 #status predicted #label CC3\
   23-655,739-1651    #product complement C3b #status predicted #label C3B\
   660-1651           #product complement C3 alpha chain #status predicted
                        #label CC3A\
   660-738            #product C3a anaphylatoxin #status predicted #label C3T\
   739-1651           #product complement C3b alpha' chain #status predicted
                        #label C3BA\
   1412-1445          #region properdin binding\
   546-807,615-650,
   683-710,684-717,
   697-718,863-1501,
   1091-1147,
   1346-1477,
   1377-1446,
   1494-1499,
   1506-1578,
   1525-1649,
   1625-1634          #disulfide_bonds #status predicted\
   738-739            #cleavage_site Arg-Ser (C3 convertase) #status
                        predicted\
   999-1002           #cross-link thiolester (Cys-Gln) #status predicted
SUMMARY         #length 1651  #molecular-weight 184925  #checksum 3526

  Query Match            19.6%;  Score 2367;  DB 2;  Length 1651;
  Best Local Similarity  30.2%;  Pred. No. 0.00e+00;
  Matches  518;  Conservative  416;  Mismatches  655;  Indels  127;  Gaps  97;

Db      4 malylvaalligfpg-s-shgalytlitpavlrtdteeqilveahgdstpksldifvhdf 61
          |:|  :  :|| | | : ::   | : :| ::|  : | |::: | : :    | : :
Qy      1 MGLLGILCFLI-FLGKTWGQEQTYVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSY 59

Db     62 prkqktlfqsrvdmnqagsmfvtpti-kvpakelnkdskqnqyvvvkvtgpqvalekvvl 120
          | |  :   ::| :: : ::| ::|  : :|:|    :    || : | : : :  | :
Qy     60 PDKKFSYSSGHVHLS-SENKFQNSAILTIQPKQLPGGQNPVSYVYLEVVSKHFSKSKRMP 118

Db    121 lsyqsgfvfiqtdkgiytpgspvryrvfsvdhnmhrmdktvivefqtpegivvs-skpvn 179
          ::|::||:||:||| :||| :|: ||:|:: ::    :  :: |  |||  |      ::
Qy    119 ITYDNGFLFIHTDKPVYTPDQSVKVRVYSLNDDLKPAKRETVLTFIDPEGSEVDMVEEID 178

Db    180 psgsir-p-ynlpelvsfgtwkavakyehspeesytayfdvreyvlpsfevrlqpsdkfl 237
          | |  | :::|    :| |   |||      : ||||:|:||||| | | ::|  :|:
Qy    179 HIGIISFPDFKIPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEPEYNFI 238

Db    238 -yidgnknfhvsitarylygkkv-egvafvvfgvk--i-ddakksipdsltripiidgdg 292
          | :  ||| ::| |||:| | | |:   :: ||::  : ||  |  :  ::    :|:| :
Qy    239 GYKN-FKNFEITIKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLINGIA 297

Db    293 eatlkrdt-lrs-rfqdlnqlvghtlyvsvtvitesgsdmvvteqggihivtspyqiyft 350
          : |:  :| ::  :  |::|    ||::||||  :|:    :|  || | ||| : :
Qy    298 QVTFDSETAVKELSYYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLNLV 357

Db    351 ktpkyfkpgmpyeltvyvtnp-d---gs-paa-hvp-v-vs-ea--ih-segtt-lsdgt 397
          || ::|||:|| : | | :: |   |: |   :   : |: |:  :  | : | : ||
Qy    358 ATPLFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDDGV 417
```

```
Db    398 akliintplniqslpitvrtnhgdlprerqaiksmtatayqtqggsenylhvaitsteik 457
          | ::|| |  :  | :.|:|:  ||| | ||  :  | || : : |  |:  :      :
Qy    418 ASFVLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKALL 477

Db    458 pgdnlpvnfnvrgnanslnqikyftylilnkgkifkvgrqprrdgqnlvtmnlhitpdli 517
          |::| : : |  ::  :: |  : ||||:||||:  | : :    :  ::|: :| :::
Qy    478 VGEHLNI-I-VTPKSPYIDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQNMV 535

Db    518 psfrfvayyqv-gnn--eivadsvwvdvkdtcmgtlvvkgassrddriqkpgaamkikle 574
          || |:: || | |:    |:|:||||::: : |    | |    |:  |    || :: :::
Qy    536 PSSRLLVYYIVTGEQTAELVSDSVWLNIEEKCGNQLQVH-LSPDADAY-SPGQTVSLNMA 593

Db    575 gdpgarvglvavdkavyvlndkykisqakiwdtieksdfgctagsgqnnlgvfedaglal 634
            ::|:| ||| ||| :     | :  :: : :||||:|| ||:| ||  ||  |||::
Qy    594 TGMDSWVALAAVDSAVYGVQRGAKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLAGLTF 653

Db    635 ttstnlntkqrsaakcpqpanrrrrssvllldskaskaaqfqdqglrkccedgmhenpmg 694
          |::| :  | :    | :    | ||:           || |:    ::||| ||    |
Qy    654 LTNANADDSQENDEPCKE-ILRPRRT---LQKKIEEIAAKYKHSVVKKCCYDGACVNNDE 709

Db    695 ytcekrakyiqegdackaaflecchyikgirdenqreselflarsdfedelfgddniisr 754
           ||| ||  |  |  | ||| |||       :| |    :| :|       |      | |
Qy    710 -TCEQRAARISLGPRCIKAFTECCVVASQLRA-NISHKDMQLGRLHMKTLLPVSKPEI-R 766

Db    755 sdfpeswlwlteeltgepnnqgisktvpfylrdsittwellavglsptkgicvaepyei 814
          | |||||||     |    |        | : | | ||:||||: ::|:| | |||||:
Qy    767 SYFPESWLWEVH-LV--PR-R----KQLQFALPDSLTTWEIQGIGISNT-GICVADTVKA 817

Db    815 tvmkdffidlrlpysvvkneqveirailynyadedi-y-vr---veliynpafcsasteg 869
           | || |::: :|||||: ||::::: :|||     : : |:   ||| |  :      :|
Qy    818 KVFKDVFLEMNIPYSVVRGEQIQLKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQG 877

Db    870 qry-r-qqfpikalssravpfvivpleqglhdveviasvrgelasdgvrkklkvvpeger 927
            :  :  :  : : ||: | | ::||| |||:::      :  :: : : | |:||||| :
Qy    878 TKSSKCVRQKVEGSSSHLVTFTVLPLEIGLHNINFSLETW--FGKEILVKTLRVVPEGVK 935

Db    928 knivtiieldpsvkgv-ggtqeltviankld-dkvpdtevetrisvlgdpvaqiiensid 985
           ::  : : ||| :|: |        :: ::  | || ||:    :|| |  |::|:   :
Qy    936 RESYSGVTLDP--RGIYGTISRRKEFPYRIPLDLVPKTEIKRILSVKGLLVGEILSAVLS 993

Db    986 gsklnhliitpsgcgeqnmitmtpsviaty-yldatgqwenlgvdrrte-a-ikqimt-g 1041
             :| |   | | : :| :::::  | |:  : ||::  :|: :  |    |    :| :  |
Qy    994 QEGINILTHLPKGSAEAELMSVVP-VFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEG 1052

Db   1042 yaqqmvykkadhsyaaftnrassswltayvvkvlamasnmvkdisheiicggvkwlilnr 1101
            | |::|| ||:      ::|:||||: ::||:  :: | : ::: || :: ||: |
Qy   1053 MLSIMSYRNADYSYSVWKGGSASTWLTAFALRVLGQVNKYV-EQNQNSICNSLLWLVENY 1111

Db   1102 qqpdgvfkenapvihgemlggtkgaep-easl--tafivtallesrsvckeqinildssi 1158
          |  :| ||||:    :  :| ||    |: | | ||  ||| | :: :  :   :|   ::| |:::
Qy   1112 QLDNGSFKENSQ-YQPIKLQGTLPVEARENSLYLTAFTVIGIRKAFDIC-PLVKI-DTAL 1168

Db   1159 nkatdyllkkyeklqrpyttaltayalaaadrlndd-r--v--lm-aa-stgrn---r-w 1207
           || ::|| :      | :| |::||||: :: : : |  |:  |   |     |   | |
Qy   1169 IKADNFLLENTLPAQSTFTLAISAYALSLGDKTHPQFRSIVSALKREALVKGNPPIYRFW 1228

Db   1208 -e--ey-narthn------iegtsyallallkmkkfaevgpvvrwlidqkyyggtyggtq 1257
           :  :  :    |      :| |:|:||: |::|  :  |  | ||::|| ::  |||  : ||
Qy   1229 KDNLQHKDSSVPNTGTARMVETTAYALLTSLNLKDINYVNPVIKWLSEEQRYGGGFYSTQ 1288

Db   1258 atvmvfqalaeyeiqmpthqdlnldisiklperevperysindrnavqartvetklnedf 1317
           |:  :::|:|| : :  :  |::|| :      :     | : |:| : :| || ||:|:
Qy   1289 DTINAIEGLTEYSL-LVKQLRLSMDIDVSYKHKGALHNYKMTDKNFL-GRPVEVLLNDDL 1346
```



```
Db   1318 tvsas-gdgkatmtiltvynaqlredanvcnkfhldvsvenvelnlkqakggkaalrlki 1376
           ||:: | | ||: : || :          :||: | | :  :::| :  :: |      |:
Qy   1347 IVSTGFGSGLATVHVTTVVH-KTSTSEEVCS-FYLKIDTQDIEASHYRGYGNSDYKRIVA 1404

Db   1377 ctry-lg--evds-tm-tiidismltgffpdaedlkrlsngvdryiskfeidnnmaqkgt 1431
           |: |   :    |   | :  :::|||: ||: :: |||| | :|||: :: ::| :: ::
Qy   1405 CASYKPSREESSSGSSHAVMDISLPTGISANEEDLKALVEGVDQLFTDYQI-KD-GH--- 1459

Db   1432 vviyldkvshsedeclhfkihkhfevgfiqpgsvkvysyynldeqctkfyhpdketglln 1491
           |:: |: :: |:  |::|:|   |||||: |::  || |    | |||:||   :     :
Qy   1460 VILQLNSIPSSDFLCVRFRIFELFEVGFLSPATFTVYEYHRPDKQCTMFYSTSNIK--IQ 1517

Db   1492 kichgnicrcaeetcsllnqq-k-kidlqlriqkacaqnvdyvyktkllrieekdgndiy 1549
           |:| |  |:| |  |: : : ::    |  : | | | |  :: | | || :  |    :    |
Qy   1518 KVCEGAACKCVEADCGQMQEELDLTISAETRKQTACKPEIAYAYKVSITSITVENVFVKY 1577

Db   1550 fmdvlevikggtdrnaqakarqyvsqrkcqealnlkldndylliwglssdlwpmk-d-dis 1607
             :|:: | |          ::      | :|: :|   :|||  |     |  :| :  :
Qy   1578 KATLLDIYKTGEAVAEKDSEITFIKKVTCTNA-ELVKGRQYLIMG-KEAL-QIKYNFSFR 1634

Db   1608 ylit-kn-twierwpnedecqd-eef-qnlcddfaq 1639
           |:    : |||| || :  |   : |   || |:||:
Qy   1635 YIYPLDSLTWIEYWPRDTTCSSCQAFLANL-DEFAE 1669
```

```
RESULT    4
ENTRY            C3HU      #type complete
TITLE            complement C3 precursor - human
CONTAINS         alternative-complement-pathway C3/C5 convertase (EC
                   3.4.21.47) C3b subunit; C3a anaphylatoxin
ORGANISM         #formal_name Homo sapiens #common_name man
DATE             28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change
                   06-Sep-1996
ACCESSIONS       A94065; A37999; A92187; A27603; A23435; A45830; B45830;
                   A01257; A01258
REFERENCE        A94065
   #authors      de Bruijn, M.H.L.; Fey, G.H.
   #journal      Proc. Natl. Acad. Sci. U.S.A. (1985) 82:708-712
   #title        Human complement component C3: cDNA coding sequence and
                   derived primary structure.
   #cross-references MUID:85140166
   #accession    A94065
      ##molecule_type mRNA
      ##residues      1-1663 ##label DEB
      ##cross-references GB:K02765
REFERENCE        A37999
   #authors      Vik, D.P.; Amiguet, P.; Moffat, G.J.; Fey, M.;
                   Amiguet-Barras, F.; Wetsel, R.A.; Tack, B.F.
   #journal      Biochemistry (1991) 30:1080-1085
   #title        Structural features of the human C3 gene: intron/exon
                   organization, transcriptional start site, and promoter
                   region sequence.
   #cross-references MUID:91113687
   #contents     intron/exon structure of gene
   #accession    A37999
      ##molecule_type DNA
      ##residues      1-25 ##label VIK
      ##cross-references GB:M63423
      ##note          the authors translated the codon GGT for residue 6 as
                        Leu, CCC for residue 7 as Leu, and AGC for residue 8
                        as Leu
REFERENCE        A92187
   #authors      Hugli, T.E.
```

```
  #journal      J. Biol. Chem. (1975) 250:8293-8301
  #title        Human anaphylatoxin (C3a) from the third component of
                  complement.
  #cross-references MUID:76069169
  #accession    A92187
    ##molecule_type protein
    ##residues      672-680,'N',682-699,'Q',701-748 ##label HUG
REFERENCE       A27603
  #authors      Daoudaki, M.E.; Becherer, J.D.; Lambris, J.D.
  #journal      J. Immunol. (1988) 140:1577-1580
  #title        A 34-amino acid peptide of the third component of complement
                  mediates properdin binding.
  #cross-references MUID:88154452
  #accession    A27603
    ##molecule_type protein
    ##residues      1409-1563 ##label DAO
REFERENCE       A23435
  #authors      Hellman, U.; Eggertsen, G.; Engstrom, A.; Sjoquist, J.
  #journal      Biochem. J. (1985) 230:353-361
  #title        Amino acid sequence of the trypsin-generated C3d fragment
                  from human complement factor C3.
  #cross-references MUID:86025442
  #accession    A23435
    ##molecule_type protein
    ##residues      1002-1012,'E',1014-1303 ##label HEL
    ##note          sequence corresponding to residues 1072-1100 was not
                      determined but was taken from de Bruijn & Fey
                      (reference A94605)
REFERENCE       A45830
  #authors      Poznansky, M.C.; Clissold, P.M.; Lachmann, P.J.
  #journal      J. Immunol. (1989) 143:1254-1258
  #title        The difference between human C3F and C3S results from a
                  single amino acid change from an asparagine to an aspartate
                  residue at position 1216 on the alpha-chain of the
                  complement component, C3.
  #accession    A45830
    ##status        not compared with conceptual translation
    ##molecule_type DNA
    ##residues      1212-1215,'N',1217-1223 ##label POZ
    ##note          this is the C3S allele
  #accession    B45830
    ##status        not compared with conceptual translation
    ##molecule_type DNA
    ##residues      1212-1223 ##label PO2
REFERENCE       S27041
  #authors      Dolmer, K.; Sottrup-Jensen, L.
  #journal      FEBS Lett. (1993) 315:85-90
  #title        Disulfide bridges in human complement component C3b.
  #contents     annotation; disulfide bonds
COMMENT    The sequence shown is the C3 fast (C3F) allele, which is found
             mainly in Caucasian populations and is associated with increased
             incidence of certain diseases.
COMMENT    Complement C3 contains two chains, formed by removal of four
             residues and linked by a disulfide bond. Its activation by a C3
             convertase, which is the central reaction in both classical and
             alternative complement pathways, releases the C3a anaphylatoxin
             from the amino end of the alpha chain and generates C3b, which
             associates with the Bb fragment of complement factor B to form
             the alternative-complement-pathway C3/C5 convertase.
COMMENT    C3a anaphylatoxin is a vasoactive peptide and a mediator of
             inflammation.
COMMENT    C3b, with its highly reactive thiol group, binds to the surface of
             foreign particles and facilitates phagocytosis. It binds to
             complement C5 and renders it susceptible to proteolysis by the
```



```
             classical-complement-pathway C3/C5 convertase. The activity of
             C3b is regulated by proteolytic cleavage involving factors H and
             I. Its degradation products can also be biologically active.
COMMENT    The major site of synthesis of this plasma protein is the liver.
GENETICS
  #gene         GDB:C3
    ##cross-references GDB:119044
  #map_position 19p13.3-19p13.3
  #note         contains 41 exons
CLASSIFICATION  #superfamily alpha-2-macroglobulin
KEYWORDS        acute phase; complement alternate pathway; complement
                  pathway; glycoprotein; hydrolase; immune response;
                  inflammation; plasma; serine proteinase; thiolester bond
FEATURE
   1-22               #domain signal sequence #status predicted #label SIG\
   23-667             #product complement C3 and C3b beta chain #status
                        predicted #label C3BB\
   23-667,672-1663    #product complement C3 #status predicted #label CC3\
   23-667,749-1663    #product C3b #status predicted #label C3B\
   672-1663           #product complement C3 alpha chain #status predicted
                        #label CC3A\
   672-748            #product C3a anaphylatoxin #status predicted #label C3T\
   749-1663           #product C3b alpha' chain #status predicted #label C3BA\
   946-1303           #product C3dk fragment #status predicted #label CDK\
   955-1303           #product C3dg fragment #status predicted #label CDG\
   955-1001           #product C3g fragment #status predicted #label C3G\
   1002-1303          #product C3d fragment #status experimental #label C3D\
   1424-1457          #region properdin binding\
   85,939             #binding_site carbohydrate (Asn) (covalent) #status
                        experimental\
   559-816,627-662,
   693-720,694-727,
   707-728,873-1513,
   1101-1158,
   1358-1489,
   1389-1458,
   1506-1511,
   1518-1590,
   1537-1661,
   1637-1646          #disulfide_bonds #status experimental\
   748-749            #cleavage_site Arg-Ser (C3 convertase) #status
                        predicted\
   954-955            #cleavage_site Arg-Glu (complement factor I) #status
                        predicted\
   1010-1013          #cross-link thiolester (Cys-Gln) #status experimental\
   1303-1304          #cleavage_site Arg-Ser (complement factor I) #status
                        predicted\
   1320-1321          #cleavage_site Arg-Ser (complement factor I) #status
                        predicted\
   1617               #binding_site carbohydrate (Asn) (covalent) #status
                        predicted
SUMMARY         #length 1663  #molecular-weight 187163  #checksum 38

  Query Match             19.6%;  Score 2361;  DB 2;  Length 1663;
  Best Local Similarity   28.4%;  Pred. No. 0.00e+00;
  Matches  489;  Conservative  458;  Mismatches  660;  Indels  115;  Gaps   91;

Db          8 sll-ll-llthlplalg-spmysiitpnilrleseetmvleahdaqgdvpvtvtvhdfpg 64
              :|| :| :|  |  : |    | | :|:|:|: : | :|::          |:::  :|
Qy          2 GLLGILCFLIFLGKTWGQEQTYVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSYPD 61

Db         65 kklvlssektvltpatnhmgnvtftipanrefksekgrnkfvtvqatfgtqvve-kvvlv 123
              ||:  ||    |::  :   :  :|| : ::: :      :| ::   : :     | : :
Qy         62 KKFSYSSGHVHLSSENKFQNSAILTIQP-KQLPGGQNPVSYVYLEVV-SKHFSKSKRMPI 119
```

```
Db    124 slqsgylfiqtdktiytpgstvlyriftvnhkllpvgrtvmvnienpegipvkqdslssq 183
Qy    120 TYDNGFLFIHTDKPVYTPDQSVKVRVYSLNDDLKPAKRETVLTFIDPEGSEV--DMVEEI 177

Db    184 nqlgvlpls-wdipelvnmgqwkirayyenspqqvfstefevkeyvlpsfeviveptekf 242
Qy    178 DHIGIISFPDFKIPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEP-E-Y 235

Db    243 yyi-yne-kglevtitarflygkkv-egtafvifgi-qd--geqrislpeslkripiedg 296
Qy    236 NFIGYKNFKNFEITIKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLING 295

Db    297 sgevvlsrkvlldgvqnlraedlvgkslyvsatvilhsgsdmvqaersgipivtspyqih 356
Qy    296 IAQVTFDSETAVKELSYYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLN 355

Db    357 ftktpkyfkpgmpfdlmvfvtnp-d---gs-pay-rvp-vav-q--ge-dtvqsltq-gd 404
Qy    356 LVATPLFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDD 415

Db    405 gvaklsinthpsqkplsitvrtkkqelseaeqatrtmqalpystvgnsnnylhlsvlrte 464
Qy    416 GVASFVLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKA 475

Db    465 lrpgetlnvnfllrmdraheakiryytylimnkgrllkagrqvrepgqdlvvlplsittd 524
Qy    476 LLVGEHLNI-IVTPKS-PYIDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQN 533

Db    525 fipsfrlvayytligasgqrevvadsvwvdvkdscvgslvvksgqsedrqpvpgqqmtlk 584
Qy    534 MVPSSRLLVYYIVTGEQTA-ELVSDSVWLNIEEKCGNQLQVHLSPDADAYS-PGQTVSLN 591

Db    585 iegdhgarvvlvavdkgvfvlnkknkltqskiwdvvekadigctpgsgkdyagvfsdagl 644
Qy    592 MATGMDSWVALAAVDSAVYGVQRGAKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLAGL 651

Db    645 tftsssgqqtaqraelqcpqpaarrrrsvqltekrmdkvg-kypkel-rkccedgmrenp 702
Qy    652 TFLTNANADDSQENDEPCKE-ILRPRRT--L-QKKIEEIAAKYKHSVVKKCCYDGACVNN 707

Db    703 mrfscqrrtrfislgeackkvfldccnyitelrrqharashlglarsnldediiaeeniv 762
Qy    708 DE-TCEQRAARISLGPRCIKAFTECCVVASQLRANISH-KDMQLGRLHMK-TLLPVSKPE 764

Db    763 srsefpeswlwnvedlkeppkngistklmniflkdsittweilavsmsdkkgicvadpfe 822
Qy    765 IRSYFPESWLWEVH-L-VPRRK----QL-QFALPDSLTTWEIQGIGISNT-GICVADTVK 816

Db    823 vtvmqdffidlrlpysvvrneqveiravlynyrqnqelkvrvellhnpafcslattkrrh 882
Qy    817 AKVFKDVFLEMNIPYSVVRGEQIQLKGTVYNYRTS-GMQFCVKMSAVEGICTSESPVIDH 875

Db    883 qqt-vt--ippk--ssls--vpyvivplktglqevevkaavyhhfisdgvrkslkvvpeg 935
Qy    876 QGTKSSKCVRQKVEGSSSHLVTFTVLPLEIGLHNINFSLETW--FGKEILVKTLRVVPEG 933

Db    936 irmnktvavrtldperl-gregvqkedip-padlsdqvpctesetril-lqgtpvaqmte 992
Qy    934 VKRESYSGV-TLDPRGIYGTISRRKE-FPYRIPL-DLVPKTEI-KRILSVKGLLVGEILS 989

Db    993 davdaerlkhlivtpsgcgeqnmigmtptviavhyldeteqwekf-g--lekrqgaleli 1049
Qy    990 AVLSQEGINILTHLPKGSAEAELMSVVPVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKL 1049
```



```
Db   1050 kkgytqqlafrqpssafaafvkrapstwltayvvkvfslavnliaidsqvlcgavkwlil 1109
Qy   1050 KEGMLSIMSYRNADYSYSVWKGGSASTWLTAFALRVLGQVNKYVEQNQNSICNSLLWLVE 1109

Db   1110 ekqkpdgvfqedapvihqemigglr-nnnekdmaltafvlislqeakdiceeqvnslpgs 1168
Qy   1110 NYQLDNGSFKENSQYQPIKLQGTLPVEARENSLYLTAFTVIGIRKAFDICP-LVK-IDTA 1167

Db   1169 itkagdfleanymnlqrsytvaiagyala--qmgr--lkg--pllnk-flttakdk-nr- 1219
Qy   1168 LIKADNFLLENTLPAQSTFTLAISAYALSLGDKTHPQFRSIVSALKREALVKGNPPIYRF 1227

Db   1220 w-ed-pgkq--l-yn-----veatsyallalqlkdfdfvppvvrwlneqryygggygst 1269
Qy   1228 WKDNLQHKDSSVPNTGTARMVETTAYALLTSLNLKDINYVNPVIKWLSEEQRYGGGFYST 1287

Db   1270 qatfmvfqalaqyqkdapdhqelnldvslqlpsrsskithrihwesasllrseetkeneg 1329
Qy   1288 QDTINAIEGLTEYSL-LVKQLRLSMDIDVSYKHKGALHNYKMT-DKNFLGRPVEVLLNDD 1345

Db   1330 ftvtae-gkgqgtlsvvtmyhakakdqltcnkfdlkvtikpapetekrpqdakntmilei 1388
Qy   1346 LIVSTGFGSGLATVHVTTVVHKTSTSEEVCS-FYLKIDTQDIEASHYRGYGNSDYKRIVA 1404

Db   1389 ctryr-g-dqdat-ms--ildismmtgfapdtddlkqlangvdryiskyeldkafsdrnt 1443
Qy   1405 CASYKPSREESSSGSSHAVMDISLPTGISANEEDLKALVEGVDQLFTDYQI-KD-G--H- 1459

Db   1444 liiyldkvshseddclafkvhqyfnveliqpgavkvyayynleesctrfyhpekedgkln 1503
Qy   1460 VILQLNSIPSSDFLCVRFRIFELFEVGFLSPATFTVYEYHRPDKQCTMFY-S-TSNIKIQ 1517

Db   1504 klcrdelcrcaeencf-iqksdd-kvtleerldkacepgvdyvyktrlvkvqlsndfdey 1561
Qy   1518 KVCEGAACKCVEADCGQMQEELDLTISAETRKQTACKPEIAYAYKVSITSITVENVFVKY 1577

Db   1562 imaieqtiksgsdevqvgqqrtfispikcrealkleekkhylmwglss-df-wgekpnls 1619
Qy   1578 KATLLDIYKTGEAVAEKDSEITFIKKVTCTNA-ELVKGRQYLIMGKEALQIKYNFSFRYI 1636

Db   1620 yiigkdtwvehwpeedecqdeenqkqcqdlgaftesmvvfgc 1661
Qy   1637 YPLDSLTWIFYWPRDTTCSSCQAF-L-ANLDEFAEDIFLNGC 1676
```

```
RESULT    5
ENTRY               I50806     #type fragment
TITLE               complement component C3 - Japanese lamprey (fragment)
ORGANISM            #formal_name Lampetra japonica #common_name Japanese lamprey
DATE                13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
                      13-Sep-1996
ACCESSIONS          I50806
REFERENCE           I50806
   #authors         Nonaka, M.; Takahashi, M.
   #journal         J. Immunol. (1992) 148:3290-3295
   #title           Complete complementary DNA sequence of the third component of
                      complement of lamprey; implication for the evolution of
                      thioester containing protein.
  #cross-references MUID:92251197
  #accession     I50806
     ##status         preliminary; translated from GB/EMBL/DDBJ
     ##molecule_type mRNA
     ##residues     1-1673 ##label NON
```

```
##cross-references GB:D10087; NID:g222887; CDS_PID:g222888
SUMMARY          #length 1673  #checksum 8983

  Query Match            19.2%;  Score 2309;  DB 15;  Length 1673;
  Best Local Similarity  29.5%;  Pred. No. 0.00e+00;
  Matches   512;  Conservative  458;  Mismatches  636;  Indels  130;  Gaps  105;

Db      1 vl-llmsvgtsvtqdpmvllsvpsviligsdvnvlvdhaastedvsvvvraeefltkk-q 58
Qy      6 ILCFLIFLGKTWGQEQTYVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSYPDKKFS 65

Db     59 latqtitltqldpaiatlklgfdienpdktnsastkhhvrlv-akvesksfnkeitahal 117
Qy     66 YSSGHVHLSS-ENKFQNSAI-LTIQ-P-KQLPGG-QNPVSYVYLEVVSKHFSKS-K-RMP 118

Db    118 lsyrsghvvvqtdkpiytpdekvkyrmfpmnredvhripvrqsmtvdivnadgviverqi 177
Qy    119 ITYDNGFLFIHTDKPVYTPDQSVKVRVYSLN-DDL-K-PAKRETVLTFIDPEGSEVDM-V 174

Db    178 ktikatdegivdgtsftipaiskhgtwkifarmsgapninssaefdvreyilptfevkin 237
Qy    175 EEI-D-HIGIISFPDFKIPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIE 232

Db    238 pkqrvfhind-eefvvditanyfnqelvsgta-yvryfl-en-gdvpk-lvdss--sttl 290
Qy    233 PEYNFIGYKNFKNFEITIKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTML 292

Db    291 vageglsilkkekllk-l-fpnakdllafsltikttvlssqaaeteeaelvgikivesry 348
Qy    293 INGIAQVTFDSETAVKELSYYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPY 352

Db    349 qitatktsryfkpelpyfiqvevrnadgspskevdvv--ak-vqvgsat--inpqk-m-r 401
Qy    353 KLNLVATPLFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTR 412

Db    402 tdsngltsftvt-ppnvnqltvtvrtderh-psneqgelvytaqkyas-a-symhid-v- 455
Qy    413 VD-DGVASFVLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTD 471

Db    456 t-rimrlgetlnvfltakttqlnavthftymvltrqvivktnrktkesgggpsnvripit 514
Qy    472 NHKALLVGEHLNIIVTPKSPYIDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVT 531

Db    515 pdmaprfrflayyilpg---geivadsvtvevtelcksqvslslkgrp-tlepkamltld 570
Qy    532 QNMVPSSRLLVYYIVTGEQTAELVSDSVWLNIEEKCGNQLQVHLSPDADAYSPGQTVSLN 591

Db    571 ligepdarvgllavdqavyavnrkhrltqdrvwkametfdtgctaeggagrpgvfsdagl 630
Qy    592 MATGMDSWVALAAVDSAVYGVQRGAKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLAGL 651

Db    631 alitskglnttdrseigcpkvpsrkprqlsmlqirreaekytqef-rkccvdglkmsptg 689
Qy    652 TFLTNANADDSQENDEPCKEIL-RPRRTLQK-KIEEIAAKYKHSVVKKCCYDGACVN-ND 708

Db    690 qgceerlkrvt-gpkecvdaflqcckkaeeyrkseslgaktvlrrndfmeldlmnedevn 748
Qy    709 ETCEQRAARISLGPR-CIKAFTECCVVASQLRANISHKDMQ-LGRLH-MK-TLLPVSKPE 764

Db    749 mmayfpqswgwnkyknsckygrhpqirlqlpdtittwnmqavsisktrgvcladplllvs 808
Qy    765 IRSYFPESWLWEVHLVP-R--RK-QLQFALPDSLTTWEIQGIGISNT-GICVADTVKAKV 819

Db    809 tkdffiklhlpysvkrgeqteirvilynymeesltiltemdivesicst-sk--s--gak 863
```



```
Qy    820 FKDVFLEMNIPYSVVRGEQIQLKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTK 879

Db    864 psq--kstvkgkgamvvsfpivplkigehhisirsrvygrtfgdgvq-kilrvapegv-r 919
Qy    880 SSKCVRQKVEGSSSHLVTFTVLPLEIGLHNINF-S-LETW-FGKEILVKTLRVVPEGVKR 936

Db    920 di-r--s-esrsvh-veeretffikneispdvvpnsdvltfisvkgdelaetmvncldak 974
Qy    937 ESYSGVTLDPRGIYGTISRRKEF-PYRIPLDLVPKTEIKRILSVKGLLVGEILSAVLSQE 995

Db    975 sisnliqiptgcgeqnmikmapttltliyldsvqeweki-g--lhrreeaigflkqgysr 1031
Qy    996 GINILTHLPKGSAEAELMSVVPVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEGMLS 1055

Db   1032 elsyrkadhsyaafikrpsstwltafvvkvyslakrviivdnqelcgpvewiiknrqnsd 1091
Qy   1056 IMSYRNADYSYSVWKGGSASTWLTAFALRVLGQVNKYVEQNQNSICNSLLWLVENYQLDN 1115

Db   1092 gsyredgpvihremqgg--vggteghvsmtafiligiqqaqeycgvsvpnykqsmnravq 1149
Qy   1116 GSFKENSQYQPIKLQGTLPVEARENSLYLTAFTVIGIRKAFDICPL-V-KIDTALIKADN 1173

Db   1150 flaskvsdlkrmytiaitryalalqd---peseaahsswkklenrttfeskghrywkae- 1205
Qy   1174 FLLENTLPAQSTFTLAISAYALSLGDKTHPQFRSIVSALKR-EALVKGNPPIYRFWKDNL 1232

Db   1206 e-t-shvlrmsais-veatayglltylrkkdyesareivdwlteqrnygggfqstqdtil 1262
Qy   1233 QHKDSSVPNTGTARMVETTAYALLTSLNLKDINYVNPVIKWLSEEQRYGGGFYSTQDTIN 1292

Db   1263 alqamaqykmdsssskel-idvqleit-spknnfekkmkiteetrfvqephkippggniti 1320
Qy   1293 AIEGLTEYSL--LVKQLRLSMDIDVSYKHKGALHN-YKMTDKN-FLGRPVEVLLNDDLIV 1348

Db   1321 kas-grgtftlsimsvfnkvapsskscstfdlkvtmteaddgespqgrlgwfdgkrrrrr 1379
Qy   1349 STGFGSGLATVHVTTVVHKTSTSEEVCS-FYLKID-TQ-DI-EASHYR-G-Y-GNSDYKR 1401

Db   1380 digdeggveavyrmnmctrykprkedlssesgmtiievnmltgfipdkndliqlkesvdk 1439
Qy   1402 -I-----V-A------CASYKPSREESSSGSSHAVMDISLPTGISANEEDLKALVEGVDQ 1448

Db   1440 yisnyeitdsvliiywdkvpstedycfafkikqmlrsdmiqpvtasvydyyspadkctrl 1499
Qy   1449 LFTDYQIKDGHVILQLNSIPSSDFLCVRFRIFELFEVGFLSPATFTVYEYHRPDKQCTMF 1508

Db   1500 ynlpggyvelsplcqndlcqcvevscpakkpkfdtsitvlhrqeaacvagidyay-vgiv 1558
Qy   1509 YS-TSN-IKIQKVCEGAACKCVEADCGQMQEELDLTISAETRKQTACKPEIAYAYKVSIT 1566

Db   1559 dnrtevgsfvyytvniqtvikegqdqaiqpkatrl-fivtrscdgrlgmetprqyllmgr 1617
Qy   1567 SITVE-NVFVKYKATLLDIYKTG-E-AVAEKDSEITFIKKVTCTNA-ELVKGRQYLIMGK 1622

Db   1618 kg-etk-drndrfqyvldasswveqwpvdekcnqpnvqtfcaikreyefsmqiqgc 1671
Qy   1623 EALQIKYNFSFRYIYPLDSLTWIEYWPRDTTCS--SCQAFLANLDEFAEDIFLNGC 1676

RESULT     6
ENTRY           I50711     #type complete
TITLE           complement C3 precursor - chicken
ORGANISM        #formal_name Gallus gallus #common_name chicken
```

```
DATE           13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
                 16-Feb-1997
ACCESSIONS     I50711
REFERENCE      I50711
   #authors    Mavroidis, M.; Sunyer, J.O.; Lambris, J.D.
   #journal    J. Immunol. (1995) 154:2164-2174
   #title      Isolation, primary structure, and evolution of the third
                 component of chicken complement and evidence for a new
                 member of the alpha 2-macroglobulin family.
   #cross-references MUID:95173423
   #accession     I50711
      ##status        preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues      1-1652 ##label MAV
      ##cross-references EMBL:U16848; NID:g755814; CDS_PID:g755815
CLASSIFICATION  #superfamily alpha-2-macroglobulin
SUMMARY         #length 1652  #molecular-weight 184085  #checksum 6420

  Query Match           19.1%;  Score 2298;  DB 6;  Length 1652;
  Best Local Similarity  29.1%;  Pred. No. 0.00e+00;
  Matches   502;  Conservative  455;  Mismatches  645;  Indels  122;  Gaps   91;

Db      1 mgll-llplllgvlllhavptpaqmvtmvtpavlrldtdekvvleapglsapteanilvq 59
          |||| :| :|  ::|  :       |     : :| ::|: : |::|::  | : : :| | :
Qy      1 MGLLGILCFL--IFLGKT-WGQEQTYVISAPKIFRVGASENIVIQVYGYTEAFDATISIK 57

Db     60 dfpqkrkvlfqvrkqlnpaegmmaiatvkvpvkllpp-vqk-hfvsvvarvgqvtlekvl 118
          :|:|:      : :|:: :       | : :  | ||          :| :      : : | :
Qy     58 SYPDKKFSYSSGHVHLSSENKFQNSAILTIQPKQLPGGQNPVSYVYLEVVSKHFSKSKRM 117

Db    119 lvslqsghiflqtdkpiytpgstvlsrlfalshfmqpllktvivevktpdnviikqvpvs 178
           :: ::| :|::|||:||| :|  |:::|:  : |  :  ::     |:    : :: |
Qy    118 PITYDNGFLFIHTDKPVYTPDQSVKVRVYSLNDDLKPAKRETVLTFIDPEGSEV-DM-VE 175

Db    179 spmrngifsi-nhnlpevvslgtwtitakfedsqdqvfstqfevkeyvlpsfevtldpqe 237
           : ||:|: : ::|     | ||| ||: :      :: |||||||| | |:::|:
Qy    176 EIDHIGIISFPDFKIPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEPEY 235

Db    238 kfl-yidpaedfrvtitarylygknl-qgtafvlfgvvvd--dekktipqs-lqrvkvtd 292
          :|: | :    :| :|| |||:| | : ::  :: ||:   |   |: | : |: :|   :: :
Qy    236 NFIGYKN-FKNFEITIKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLIN 294

Db    293 gdgqavl-pmamlrq-pfanlqelvghslyvtvtvltesgsdmveaqrsgirivtspyti 350
          | :|  : : : ::: :: :|::|      ||::|||:  :|:    ||: :||: | ||| :
Qy    295 GIAQVTFDSETAVKELSYYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKL 354

Db    351 hfthtpkyfkpgmpfdptvyvtnp-dn---s-pa--aa-gipvkad--nfqglvstqr-- 398
          ::  || ::|||:|:    | | :: |      : |        | | |: :  :::    |  |
Qy    355 NLVATPLFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVD 414

Db    399 dgtaklvlnmpanknsvpitvrtdqkdlpperqasrqivaeayqsgqnsgnylhlavgas 458
          || | :|||:|:      : : |:|| ||| | ||       | ||| | : |  |:  :
Qy    415 DGVASFVLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHK 474

Db    459 qvqpgdnlpinfhlksnrddvrksvsyftylilskghivhvgrqpregdqslvtmslpvt 518
           :  |::| | : : ::   : | :: : |||||| |:| | : : :| |   ::::|||
Qy    475 ALLVGEHLNI-I-VTPKSPYIDK-ITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVT 531

Db    519 anlipsfrivayyhvk---pgeiiadsvwvdvkdtcmgslvvrgaseadnrvheprtpmr 575
          |::|| |:: || |      :|:::||||::: : |   | |: :::||      |     :
Qy    532 QNMVPSSRLLVYYIVTGEQTAELVSDSVWLNIEEKCGNQLQVHLSPDADA--YSPGQTVS 589

Db    576 lhiegdhkahvglvavdkavyvlnkn-kltqskvwdtvensdigctpgsgrnqvgvfada 634
          |::       : |:| ||| ||| : :  |      :| : :|:||:|| :|:| |     || |
```



```
Qy    590 LNMATGMDSWVALAAVDSAVYGVQRGAKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLA 649

Db    635 glsltsnvninteqrsevqcakpakrkrrsvrlikhkgtkmaeysdknlrkccedgirkn 694
          ||:: :| | :  | ::  | |   | ||:  | |       | |      ::||| ||   |
Qy    650 GLTFLTNANADDSQENDEPC-KEILRPRRT--LQKKIEEIAAKYKHSVVKKCCYDGACVN 706

Db    695 lmgyscekratyv-ldakscteaflscclyikgirdeerelqyelarsevddaflsdedi 753
             :|| ||: : | :: |  ||  ||:     :| :      : :|:|  :  ::|:
Qy    707 -NDETCEQRAARISLGPR-CIKAFTECCVVASQLRANISHKDMQLGRLHMK-TLLPVSKP 763

Db    754 tsrslfpeswlwqveelteppneqqismktlpiylkdsittwevlavsisenkglcvadp 813
            || |||||||:|  |   |    | :   :||     ||:||||: :::|| | |:||||
Qy    764 EIRSYFPESWLWEVH-LV-PRRKQ-LQF-ALP----DSLTTWEIQGIGIS-NTGICVADT 814

Db    814 yeitvmkeffidlrlpysavrneqvevrailynywtnkikvrvelmynpalcsa-stt-- 870
             | |: |::: :||| || ||::::: :|||:|: :    | :     ::|:: |
Qy    815 VKAKVFKDVFLEMNIPYSVVRGEQIQLKGTVYNYRTSGMQFCVKMSAVEGICTSESPVID 874

Db    871 -k-tryqq-ifq-lepqssdavpfvivplelgqhdvevkaavwnsfvsdgvkkklrvvpe 926
             |:     : | :|  ||   | | ::|||:| |:::       |   |   : : | ||||||
Qy    875 HQGTKSSKCVRQKVEGSSSHLVTFTVLPLEIGLHNINFSLETW--FGKEILVKTLRVVPE 932

Db    927 gmrlektvkiveldpktl-gnngvq-evkvkaanlsdivpntesetkvsiqqnpvs-ilv 983
          |:: | :    | |||: : |   : : |     : : | |:||:||     :|: |   |: ||
Qy    933 GVKRE-SYSGVTLDPRGIYGTISRRKEFPYRIP-L-DLVPKTEIKRILSVKGLLVGEILS 989

Db    984 ekatdgtklkhlivtpsgcgeqnmigmtptviavhyldstmqwetfginrrteaielikk 1043
           ::     :: |      | | :| :::: |          |||:: :|: | :   |   | ||
Qy    990 AVLSQEG-INILTHLPKGSAEAELMSVVPVFYVFHYLETGNHWNIFHSDPLIEKQKLKKK 1048

Db   1044 ---gytqqlayrkedgsfaafttrpsstwltayvakvfamainmvdikpevvcgaikwli 1100
                |     :: ||: | :::    : ::||||||: :|::    : |: : : :| :: ||:
Qy   1049 LKEGMLSIMSYRNADYSYSVWKGGSASTWLTAFALRVLGQVNKYVEQNQNSICNSLLWLV 1108

Db   1101 lekqqpdglfqedapvihkemvggyh-ga-epsvsltafvlsalqesqkicknyvksldg 1158
           : |  :| | | |::          : |       | | |: |||| : ::: :   ||   || :|
Qy   1109 ENYQLDNGSFKENSQYQPIKLQGTLPVEARENSLYLTAFTVIGIRKAFDIC-PLVK-IDT 1166

Db   1159 siakasdylsrkyqsltrpytvaltsyalaltgklnse-k--v--lmk--fsk-dgt--h 1208
          :: || ::|   :   :   :|:|:::||:| | ::: :  |  |::  : | :    :
Qy   1167 ALIKADNFLLENTLPAQSTFTLAISAYALSLGDKTHPQFRSIVSALKREALVKGNPPIYR 1226

Db   1209 -w--a-e-rna---ht-yn--iegtsyalvallqmekaeltgpvvrwlaqqnyfgggygs 1257
           |    : :::    :|      :| |:|||:: | :     :   ||::||::: :|||: |
Qy   1227 FWKDNLQHKDSSVPNTGTARMVETTAYALLTSLNLKDINYVNPVIKWLSEEQRYGGGFYS 1286

Db   1258 tqatilvfqalaqyhvalprhvelnldvsvllprranaityriennnalvarsaetklne 1317
          || ||   :::|::| : | ::: |::|: |      :::      |:: ::| | :|: |   ||:
Qy   1287 TQDTINAIEGLTEYSL-LVKQLRLSMDIDVSYKHKGALHNYKMTDKNFL-GRPVEVLLND 1344

Db   1318 dftvkae-qtgkgtmtvvtvykakvpekenkcdnfdlrvsvedvkagreve-gvirsvki 1375
          |: | :  |:| :|: | ||    | :   |: | :| |::   :|: |::      |      :|
Qy   1345 DLIVSTGFGSGLATVHVTTVVH-KTSTSEEVC-SFYLKIDTQDIEASHYRGYGNSDYKRI 1402

Db   1376 tictrf-l---dtvdatm-sildismltafspdvqdlkslseqveryiskfeidhalsnr 1430
           |: :     ::  ::  :::|||: |::|:: :|||:| ||||:: :   |   |   : :
Qy   1403 VACASYKPSREESSSGSSHAVMDISLPTGISANEEDLKALVEGVDQ-L--F-TDYQIKD- 1457

Db   1431 snliiyldkvshqveeciafrahqhfqvgliqpasvivysyykiddrctrfyhpdkaggq 1490
          :::|: |: ::       |: ||   : |:||:: ||:  || | : | :|| ||    :
Qy   1458 GHVILQLNSIPSSDFLCVRFRIFELFEVGFLSPATFTVYEYHRPDKQCTMFYSTSNI--K 1515

Db   1491 lrkichgevc-caeencf-irvkkdnpitvneridlackpgvdyvykvkvvateetpshd 1548
          ::|:| |   | | | :|   ::     |   |: : | : |||| : | ||| : :
```

```
Qy    1516 IQKVCEGAACKCVEADCGQMQEELDLTISAETRKQTACKPEIAYAYKVSITSITVENVFV 1575

Db    1549 nyimsiltvikmgtdenpggsnrtfvshkqcrdalslqkgqdylvwglas-dl-wvtgsr 1606
           :|  ::| : | |       |: ||:    | :|  | ||::||: |  : ::    : |
Qy    1576 KYKATLLDIYKTGEAVAEKDSEITFIKKVTCTNA-ELVKGRQYLIMGKEALQIKYNFSFR 1634

Db    1607 fsyliskdtwleawpleescqdadlqplcqdftefsdnlvlfgc 1650
           : | :   ||:| || : :|    :    |   :: ||:::: | ||
Qy    1635 YIYPLDSLTWIEYWPRDTTCSSCQ-AFL-ANLDEFAEDIFLNGC 1676


RESULT     7
ENTRY           C3MS        #type complete
TITLE           complement C3 precursor - mouse
CONTAINS        alternative-complement-pathway C3/C5 convertase (EC
                  3.4.21.47) C3b subunit; C3a anaphylatoxin
ORGANISM        #formal_name Mus musculus #common_name house mouse
DATE            30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change
                  06-Sep-1996
ACCESSIONS      A92459; B92459; A92460; A93938; A21898; A54561; S16369;
                  S16189; I49563; I49576; A01261; A05290; A29033
REFERENCE       A92459
   #authors     Lundwall, A.; Wetsel, R.A.; Domdey, H.; Tack, B.F.; Fey, G.H.
   #journal     J. Biol. Chem. (1984) 259:13851-13856
   #title       Structure of murine complement component C3: I. Nucleotide
                  sequence of cloned complementary and genomic DNA coding for
                  the beta chain.
   #cross-references MUID:85054818
   #accession   A92459
      ##molecule_type mRNA
      ##residues     1-724 ##label LU1
   #accession   B92459
      ##molecule_type DNA
      ##residues     1-124 ##label LU2
REFERENCE       A92460
   #authors     Wetsel, R.A.; Lundwall, A.; Davidson, F.; Gibson, T.; Tack,
                  B.F.; Fey, G.H.
   #journal     J. Biol. Chem. (1984) 259:13857-13862
   #title       Structure of murine complement component C3: II. Nucleotide
                  sequence of cloned complementary DNA coding for the alpha
                  chain.
   #cross-references MUID:85054819
   #accession   A92460
      ##molecule_type mRNA
      ##residues     671-1663 ##label WET
REFERENCE       A93938
   #authors     Domdey, H.; Wiebauer, K.; Kazmaier, M.; Muller, V.; Odink,
                  K.; Fey, G.
   #journal     Proc. Natl. Acad. Sci. U.S.A. (1982) 79:7619-7623
   #title       Characterization of the mRNA and cloned cDNA specifying the
                  third component of mouse complement.
   #cross-references MUID:83117730
   #contents    C3a
   #accession   A93938
      ##molecule_type mRNA
      ##residues     671-748 ##label DOM
REFERENCE       A21898
   #authors     Sottrup-Jensen, L.; Stepanik, T.M.; Kristensen, T.; Lonblad,
                  P.B.; Jones, C.M.; Wierzbicki, D.M.; Magnusson, S.; Domdey,
                  H.; Wetsel, R.A.; Lundwall, A.; Tack, B.F.; Fey, G.H.
   #journal     Proc. Natl. Acad. Sci. U.S.A. (1985) 82:9-13
   #title       Common evolutionary origin of alpha2-macroglobulin and
                  complement components C3 and C4.
   #cross-references MUID:85113177
```



```
   #accession   A21898
      ##molecule_type mRNA
      ##residues     25-1663 ##label SOT
REFERENCE       A54561
   #authors     Hamada, J.; Cavanaugh, P.G.; Miki, K.; Nicolson, G.L.
   #journal     Cancer Res. (1993) 53:4418-4423
   #title       A paracrine migration-stimulating factor for metastatic tumor
                  cells secreted by mouse hepatic sinusoidal endothelial
                  cells: identification as complement component C3b.
   #accession   A54561
      ##molecule_type protein
      ##residues     25-41;749-760 ##label HAM
      ##experimental_source migration-stimulating factor purified from medium
                      conditioned by mouse hepatic sinusoidal
                      endothelial cells
REFERENCE       S16189
   #authors     Sato, T.; Hong, M.H.; Jin, C.H.; Ishimi, Y.; Udagawa, N.;
                  Shinki, T.; Abe, E.; Suda, T.
   #journal     FEBS Lett. (1991) 285:21-24
   #title       The specific production of the third component of complement
                  by osteoblastic cells treated with 1-alpha,
                  25-dihydroxyvitamin D(3).
   #cross-references MUID:91293304
   #accession   S16369
      ##molecule_type protein
      ##residues     25-31 ##label SAT
   #accession   S16189
      ##status       preliminary
      ##molecule_type protein
      ##residues     671-677,'X',679-680 ##label SA2
REFERENCE       I49563
   #authors     Fey, G.; Domdey, H.; Wiebauer, K.; Whitehead, A.S.; Odink, K.
   #journal     Springer Semin. Immunopathol. (1983) 6:119-147
   #title       Structure and expression of the C3 gene.
   #cross-references MUID:84045280
   #accession   I49563
      ##status       preliminary
      ##molecule_type mRNA
      ##residues     25-136,'Q',138-240 ##label FEY
      ##cross-references GB:M35659; NID:g192280; CDS_PID:g192281
REFERENCE       I49576
   #authors     Fey, G.H.; Wiebauer, K.; Domdey, H.
   #journal     Ann. N. Y. Acad. Sci. (1983) 421:307-312
   #title       Amino acid sequences of mouse complement C3 derived from
                  nucleotide sequences of cloned cDNA.
   #cross-references MUID:84201365
   #accession   I49576
      ##status       preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues     658-761 ##label RES
      ##cross-references GB:M33032; NID:g192391; CDS_PID:g192392
COMMENT    Complement C3 contains two chains, formed by removal of four
             residues and linked by a disulfide bond. Its activation by a C3
             convertase, which is the central reaction in both classical and
             alternative complement pathways, releases the C3a anaphylatoxin
             from the amino end of the alpha chain and generates C3b, which
             associates with the Bb fragment of complement factor B to form
             the alternative-complement-pathway C3/C5 convertase.
COMMENT    C3a anaphylatoxin is a vasoactive peptide and a mediator of
             inflammation.
COMMENT    C3b, with its highly reactive thiol group, binds to the surface of
             foreign particles and facilitates phagocytosis. It binds to
             complement C5 and renders it susceptible to proteolysis by the
             classical-complement-pathway C3/C5 convertase. The activity of
```

```
          C3b is regulated by proteolytic cleavage involving factors H and
          I. Its degradation products can also be biologically active.
COMMENT    The major site of synthesis of this plasma protein is the liver.
GENETICS
  #introns        27/2; 90/3
  #note           the list of introns may be incomplete
CLASSIFICATION  #superfamily alpha-2-macroglobulin
KEYWORDS        acute phase; complement alternate pathway; complement
                  pathway; glycoprotein; hydrolase; immune response;
                  inflammation; plasma; serine proteinase; thiolester bond
FEATURE
  1-24                #domain signal sequence #status predicted #label SIG\
  25-666              #product complement C3 and C3b beta chain #status
                        predicted #label C3BB\
  25-666,671-1663     #product complement C3 #status predicted #label CC3\
  25-666,749-1663     #product C3b #status predicted #label C3B\
  671-1663            #product complement C3 alpha chain #status predicted
                        #label CC3A\
  671-748             #product C3a anaphylatoxin #status predicted #label C3T\
  749-1663            #product C3b alpha' chain #status predicted #label C3BA\
  946-1303            #product C3dk fragment #status predicted #label CDK\
  1002-1303           #product C3d fragment #status predicted #label C3D\
  1424-1457           #region properdin binding\
  559-816,626-661,
  693-720,694-727,
  707-728,873-1513,
  1101-1158,
  1358-1489,
  1389-1458,
  1506-1511,
  1518-1590,
  1537-1661,
  1637-1646           #disulfide_bonds #status predicted\
  748-749             #cleavage_site Arg-Ser (C3 convertase) #status
                        predicted\
  939,1617            #binding_site carbohydrate (Asn) (covalent) #status
                        predicted\
  1010-1013           #cross-link thiolester (Cys-Gln) #status predicted\
  1303-1304           #cleavage_site Arg-Ser (complement factor I) #status
                        predicted\
  1320-1321           #cleavage_site Arg-Ser (complement factor I) #status
                        predicted
SUMMARY          #length 1663  #molecular-weight 186482  #checksum 646

  Query Match          19.1%;  Score 2296;  DB 2;  Length 1663;
  Best Local Similarity   28.2%;  Pred. No. 0.00e+00;
  Matches  479;  Conservative  447;  Mismatches  665;  Indels  106;  Gaps   84;

Db      28 ysiitpnvlrleseetivleahdaqgdipvtvtvqdflkrqvltsektvltgasghlrsv 87
              | | :|:::|: : | ||::       : |::: :  :    |    |  ::   :::
Qy      23 YVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSYPDKKFSYSSGHVHLSSENKFQNS 82

Db      88 sikipaskefnsdkeghkyvtvvanfqetvvekavmvsfqsgylfiqtdktiytpgstvl 147
              :|   :|:: :  :   ||  :             | : ::::|:|||:||| :||| :|
Qy      83 AILTIQPKQLPGGQNPVSYVYLEVVSKHFSKSKRMPITYDNGFLFIHTDKPVYTPDQSVK 142

Db     148 yriftvdnnllpvgk-tvvilietpdgipvkrdilssnnqhgilpls-wnipelvnmgqw 205
              |:::::::| | : ||: :|: |:| |  |::   :: ||:::: :||      | |
Qy     143 VRVYSLNDDLKPAKRETVLTFID-PEGSEV--DMVEEIDHIGIISFPDFKIPSNPRYGMW 199

Db     206 kirafyehapkqifsaefevkeyvlpsfevrveptetfy-yiddpnglevsiiakflygk 264
              |:| |        :| |||||||| | | :||   |  |  : : :|::| |:::| |
Qy     200 TIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEPEYNFIGY-KNFKNFEITIKARYFYNK 258
```



```
Db     265 nv-dgtafvifgv-qd--gdkkislahsltrvviedgvgdavltrkvlmegvrpsnadal 320
             | ::  :: ||: :|  | |  :  ::   :: :|:::  :      :  :   : : |
Qy     259 VVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLINGIAQVTFDSETAVKELSYYSLEDL 318

Db     321 vgksly vsvtvilhsgsdmveaersgipivtspyqihftktpkffkpampfdlmvfvtnp 380
             | ||::|||| :|:   ||| :||  | ||| :::  || |:||::|: ::| | ::
Qy     319 NNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLNLVATPLFLKPGIPYPIKVQVKDS 378

Db     381 -d---gs-pas-kvl-v-vtqg-sn---akaltq-ddgvaklsintpnsrqpltitvrtk 427
             |   |: |   :     : | |  |:   :|::|: |||| : :| |::   | : |:|
Qy     379 LDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDDGVASFVLNLPSGVTVLEFNVKTD 438

Db     428 kdtlpesrqatktmeahpystmhnsnnylhlsvsrmelkpgdnlnvnfhlrtdpgheaki 487
              ||| ||      | :||::  |   |:  : ::: |  |::||:     :: |     ||
Qy     439 APDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKALLVGEHLNIIVTPKS-P-YIDKI 496

Db     488 ryytylvmnkgkllkagrqvrepgqdlvvlslpitpefipsfrlvayytligasgqrevv 547
             | ||:::|||::  | : : :        :::|:| : :|| ||: || : |      |:|
Qy     497 THYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQNMVPSSRLLVYYIVTGEQTA-ELV 555

Db     548 adsvwvdvkdscigtlvvkgdprdnhlapgqqttlriegnqgarvglvavdkgvfvlnkk 607
             :||||::: : |    | |    |  :  :|||  :| :      ::|:| ||| :|: : :
Qy     556 SDSVWLNIEEKCGNQLQVHLSPDADAYSPGQTVSLNMATGMDSWVALAAVDSAVYGVQRG 615

Db     608 nkltqskiwdvvekadigctpgsgknyagvfmdaglafktsqglqteqradlectkpaar 667
             |     :: : :||:|:|| :|:| | | ||  |||:| |:    :  |  |  | |   |
Qy     616 AKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLAGLTFLTNANADDSQENDEPC-KEILR 674

Db     668 rrrsvqlmerrmdkagqqytdkglrkccedgmrdipmryscqrrarlitqgencikafidc 727
             ||::| : :  : |: |     ::||| ||    :   :|::||  |: |  ||||| :|
Qy     675 PRRTLQ-K-KIEEIAAKYKHSVVKKCCYDGAC-VNNDETCEQRAARISLGPRCIKAFTEC 731

Db     728 cnhitklreqhrrdhvlglarseleedi ipeediisrshfpqswlwtieelkepekngis 787
             |   : ||    :   : |:|  :     ::|       || ||:|||| :  |  | :
Qy     732 CVVASQLRANISHKD-MQLGRLHMK-TLLPVSKPEIRSYFPESWLWEVH-L-VP-RR--- 783

Db     788 tkvmniflkdsittweilavslsdkkgicvadpyeirvmqdffidlrlpysvvrneqvei 847
             | : : | ||:||||| ::::|:  |||||      :|   | |::: :|||||| ||:::
Qy     784 -KQLQFALPDSLTTWEIQGIGISNT-GICVADTVKAKVFKDVFLEMNIPYSVVRGEQIQL 841

Db     848 ravlfnyreqe-el--kvr-ve-l-l-hnpafcsmataknryfqtikippkssvavpyvi 900
             :: ::|||    ::  |:   ||  :    :| :    :|  ::   :   |:    || | : :
Qy     842 KGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTKSSKCVRQ-KVEGSSSHLVTFTV 900

Db     901 vplkiqqqevevkaavfnhfisdgvkktlkvvpegmrinktvaihtldpeklgqggv-qk 959
             :|| || ::::           |   : : |||:||||||:: :   :: ||||  :   | : ::
Qy     901 LPLEIGLHNINFSLETW--FGKEILVKTLRVVPEGVKRESYSGV-TLDPRGI-YGTISRR 956

Db     960 vdvp-aadlsdqvpdtdsetrii-lqgspvvqmaedavdgerlkhlivtpagcgeqnmig 1017
               :    | | | |:    ||: :   |   :        :  |  : : |    | | :| ::::
Qy     957 KEFPYRIPL-DLVPKTEI-KRILSVKGLLVGEILSAVLSQEGINILTHLPKGSAEAELMS 1014

Db    1018 mtptviavhyldqteqwekfgiek-rq-ealel-ikkgytqqlafkqpssayaafnnrpp 1074
             : |             |||:  ::|: :   : : |  :  ::::: :  :|:   :  ::
Qy    1015 VVPVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEGMLSIMSYRNADYSYSVWKGGSA 1074

Db    1075 stwltayvvkvfslaanliaidshvlcgavkwlilekqkpdgvfqedgpvihqemiggfr 1134
             ||||||: ::|::   : :   : : : :| :: ||: : |   :| | |::       : | :
Qy    1075 STWLTAFALRVLGQVNKYVEQNQNSICNSLLWLVENYQLDNGSFKENSQYQPIKLQGTLP 1134

Db    1135 -nakeadvsltafvlialqeardicegqvnslpgsinkageyieasymnlqrpytvaiag 1193
             :|:| : |||| :|::: | |||    |: :  :: || :::  : :  |  :|:||::
Qy    1135 VEARENSLYLTAFTVIGIRKAFDICP-LVK-IDTALIKADNFLLENTLPAQSTFTLAISA 1192
```

```
Db   1194 yalalmnkleepy--lgkflnt-a--k-drn--r-w-e--ep-dqql-yn-----veats 1234
          |||:| :|     :  :   |:  |  | :     | | :  :  |  :       ||:|:
Qy   1193 YALSLGDKTHPQFRSIVSALKREALVKGNPPIYRFWKDNLQHKDSSVPNTGTARMVETTA 1252

Db   1235 yallalllkdfdsvppvvrwlneqryygggygstqatfmvfqalaqyqtdvpdhkdlnm 1294
          ||||: | |||:: | ||::||:|:: ||||: ||| |:  :::|::|   | : : |:|
Qy   1253 YALLTSLNLKDINYVNPVIKWLSEEQRYGGGFYSTQDTINAIEGLTEYSLLVKQLR-LSM 1311

Db   1295 dvsfhlpsrssattfrllwengnllrseetkqneafsltak-gkgrgtlsvvavyhaklk 1353
          |:      :::   :::  ::  | |: |    |: : :::  | | :|: | :| |
Qy   1312 DIDVSYKHKGALHNYKMT-DKNFLGRPVEVLLNDDLIVSTGFGSGLATVHVTTVVHKTST 1370

Db   1354 skvtckkfdlrvsirpapetakkpeeakntmfleictky-lg-dvdat-ms--ildismm 1408
          |   |  | |::  :     :  :        : : :  |: |  : :  ::  |  ::|||:
Qy   1371 SEEVCS-FYLKIDTQDIEASHYRGYGNSDYKRIVACASYKPSREESSSGSSHAVMDISLP 1429

Db   1409 tgfapdtkdlellasgvdryiskyemnkafsnkntliiylekishteedcltfkvhqyfn 1468
          ||::::  ||  |  |||: :: |:: |  ::    :|: |: |: ::  |: |:: : |:
Qy   1430 TGISANEEDLKALVEGVDQLFTDYQI-KD-GH---VILQLNSIPSSDFLCVRFRIFELFE 1484

Db   1469 vgliqpgsvkvysyynleesctrfyhpekddgmlsklchsemcrcaeencf-mqqsqe-k 1526
          ||:: |::  || |    :  || || :    : :: |:| :  |:| | :|  ||:  :
Qy   1485 VGFLSPATFTVYEYHRPDKQCTMFY-S-TSNIKIQKVCEGAACKCVEADCGQMQEELDLT 1542

Db   1527 inlnvrldkacepgvdyvykteltnikllddfdeytmtiqqviksgsdevqagqqrkfis 1586
          |: : | : || | : | ||  :|:| : : |  |  |: :: |:|     :   :  ||
Qy   1543 ISAETRKQTACKPEIAYAYKVSITSITVENVFVKYKATLLDIYKTGEAVAEKDSEITFIK 1602

Db   1587 hikcrnalklqkgkkylmwglss-dl-wgekpntsyiigkdtwvehwpeaeecqdqkyqk 1644
           : | ||  | ||: ||: |  : ::              | :   ||:| ||    |    |
Qy   1603 KVTCTNA-ELVKGRQYLIMGKEALQIKYNFSFRYIYPLDSLTWIEYWPRDTTC-S-SCQA 1659

Db   1645 qceelgaftesmvvygc 1661
             :|  |:| : : ||
Qy   1660 FLANLDEFAEDIFLNGC 1676
```

```
RESULT    8
ENTRY             C3GP        #type complete
TITLE             complement C3 precursor - guinea pig
CONTAINS          alternative-complement-pathway C3/C5 convertase (EC
                    3.4.21.47) C3b subunit; C3a anaphylatoxin
ORGANISM          #formal_name Cavia porcellus #common_name guinea pig
DATE              07-Feb-1992 #sequence_revision 07-Oct-1994 #text_change
                    16-Feb-1997
ACCESSIONS        A37156; S03375; A20342; D20342; C20342; A31222
REFERENCE         A37156
   #authors       Auerbach, H.S.; Burger, R.; Dodds, A.; Colten, H.R.
   #journal       J. Clin. Invest. (1990) 86:96-106
   #title         Molecular basis of complement C3 deficiency in guinea pigs.
   #cross-references MUID:90307998
   #accession     A37156
      ##molecule_type mRNA
      ##residues      1-1666 ##label AUE
      ##cross-references GB:M34054
REFERENCE         S03375
   #authors       Gerard, N.P.; Lively, M.O.; Gerard, C.
   #journal       Protein Seq. Data Anal. (1988) 1:473-478
   #title         Amino acid sequence of guinea pig C3a anaphylatoxin.
   #cross-references MUID:89113342
   #accession     S03375
      ##molecule_type protein
      ##residues      676-730,'N',732-752 ##label GER
      ##experimental_source complement-activated guinea pig serum
```



```
      ##note          form isolated is inactive C3a anaphylatoxin and is
                        missing the carboxyl-terminal arginine
REFERENCE         A90479
   #authors       Thomas, M.L.; Tack, B.F.
   #journal       Biochemistry (1983) 22:942-947
   #title         Identification and alignment of a thiol ester site in the
                    third component of guinea pig complement.
   #cross-references MUID:83178889
   #accession     A20342
      ##molecule_type protein
      ##residues      676-687 ##label TH1
   #accession     D20342
      ##molecule_type protein
      ##residues      993-1012,1014-1017,'E',1019-1030,'Y' ##label TH2
REFERENCE         A20342
   #authors       Goldberger, G.; Thomas, M.L.; Tack, B.F.; Williams, J.;
                    Colten, H.R.; Abraham, G.N.
   #journal       J. Biol. Chem. (1981) 256:12617-12619
   #title         NH2-terminal structure and cleavage of guinea pig pro-C3, the
                    precursor of the third complement component.
   #cross-references MUID:82075767
   #accession     C20342
      ##molecule_type protein
      ##residues      23-38 ##label GOL
COMMENT     Complement C3 contains two chains, formed by removal of four
              residues and linked by a disulfide bond. Its activation by a C3
              convertase, which is the central reaction in both classical and
              alternative complement pathways, releases the C3a anaphylatoxin
              from the amino end of the alpha chain and generates C3b, which
              associates with the Bb fragment of complement factor B to form
              the alternative-complement-pathway C3/C5 convertase.
COMMENT     C3a anaphylatoxin is a vasoactive peptide and a mediator of
              inflammation.
COMMENT     C3b, with its highly reactive thiol group, binds to the surface of
              foreign particles and facilitates phagocytosis. It binds to
              complement C5 and renders it susceptible to proteolysis by the
              classical-complement-pathway C3/C5 convertase. The activity of
              C3b is regulated by proteolytic cleavage involving factors H and
              I. Its degradation products can also be biologically active.
COMMENT     The major site of synthesis of this plasma protein is the liver.
CLASSIFICATION    #superfamily alpha-2-macroglobulin
KEYWORDS          acute phase; complement alternate pathway; complement
                    pathway; glycoprotein; hydrolase; immune response;
                    inflammation; liver; plasma; serine proteinase; thiolester
                    bond
FEATURE
   1-22               #domain signal sequence #status predicted #label SIG\
   23-671             #product complement C3 and C3b beta chain #status
                        predicted #label C3BB\
   23-671,676-1666    #product complement C3 #status predicted #label CC3\
   23-671,754-1666    #product complement C3b #status predicted #label C3B\
   676-1666           #product complement C3 alpha chain #status predicted
                        #label CC3A\
   676-753            #product C3a anaphylatoxin #status predicted #label C3T\
   754-1666           #product complement C3b alpha' chain #status predicted
                        #label C3BA\
   951-1308           #product C3dk fragment #status predicted #label CDK\
   1007-1308          #product C3d fragment #status predicted #label C3D\
   1429-1461          #region properdin binding\
   557-821,630-666,
   698-725,699-732,
   712-733,878-1517,
   1106-1163,
   1363-1493,
```

```
    1394-1462,
    1510-1515,
    1522-1593,
    1540-1664,
    1640-1649          #disulfide_bonds #status predicted\
    753-754            #cleavage_site Arg-Ser (C3 convertase) #status
                         predicted\
    944,1620           #binding_site carbohydrate (Asn) (covalent) #status
                         predicted\
    1015-1018          #cross-link thiolester (Cys-Gln) #status experimental\
    1308-1309          #cleavage_site Arg-Ser (complement factor I) #status
                         predicted\
    1325-1326          #cleavage_site Arg-Ser (complement factor I) #status
                         predicted
SUMMARY         #length 1666  #molecular-weight 186487  #checksum 5419

  Query Match            19.1%;  Score 2302;  DB 2;  Length 1666;
  Best Local Similarity  28.5%;  Pred. No. 0.00e+00;
  Matches  486;  Conservative  450;  Mismatches 658;  Indels 111;  Gaps  83;

Db     23 dpmysiitpnilrleneetvvleahevqgdipvtvtvhdfpakknvlssektvltsatgy 82
Qy     20 EQTYVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSYPDKKFSYSSGHVHLSSENKF 79

Db     83 lgtvtikipaskefksdkgr-klvvvqaafggtqlekvvlvslqsgylfiqtdktiytpg 141
Qy     80 QNSAILTIQP-KQLPGGQNPVSYVYLEVVSKHFSKSKRMPITYDNGFLFIHTDKPVYTPD 138

Db    142 stvlyriftvdsdllpvgrtiivtietpdgipikrdtlssnnqhgilpls-wnipelvnm 200
Qy    139 QSVKVRVYSLNDDLKPAKRETVLTFIDPEGSEV--DMVEEIDHIGIISFPDFKIPSNPRY 196

Db    201 gqwkiqafyenspkqvfsaefevkeyvlpsfevlveptekfy-yiddpkglevniiarfl 259
Qy    197 GMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEPEYNFIGY-KNFKNFEITIKARYF 255

Db    260 ygknv-dgtafvifgv-qd--gdqrislaqsltrvviedgsgevvlsrqvlldgvqpsrp 315
Qy    256 YNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLINGIAQVTFDSETAVKELSYYSL 315

Db    316 ealvgkslyvsvtvilhsgsdmveaersgipivtspyqihftktpkyfkpampfeimvlv 375
Qy    316 EDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLNLVATPLFLKPGIPYPIKVQV 375

Db    376 tnp-d---gs-pap-hvp-v-vtqg-snvq---sltqad-gvarlsintpntrqplsvtv 422
Qy    376 KDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDDGVASFVLNLPSGVTVLEFNV 435

Db    423 qtkkggipdarqaintmqalpyttmynsnnylhlsmprtelkpgetinvnfhlrsdpnqe 482
Qy    436 KTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKALLVGEHLNIIVTPKS-PYID 494

Db    483 akiryytylimnkgkllkvgrqprepgqalvvlpmpitkelipsfrlvayytligasaqr 542
Qy    495 -KITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQNMVPSSRLLVYYIVTGEQTA- 552

Db    543 evvadsvwadvrdscvgtlvvkggsgkdgqdkrqhlprqqmtlriegnqgarvglvavd 602
Qy    553 ELVSDSVWLNIEEKCGNQLQVHLSP--DA-D---AYSPGQTVSLNMATGMDSWVALAAVD 606

Db    603 kgvfvlnkkhkltqskiwdvvekadigctpgsgkdyagvftdaglsfksskaglqtaqre 662
Qy    607 SAVYGVQRGAKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLAGLTFLTN-ANADDSQEN 665
```



```
Db    663 gldcpkpaarrrrsvqlmerrmdkagkykskelrrccedgmrenpmqfscqrraryvslg 722
Qy    666 DEPC-KEILRPRRTLQ-K-KIEEIAAKYKHSVVKKCCYDGACVNNDE-TCEQRAARISLG 721

Db    723 eacvkafldcctymaqlrqqhrreqnlglarsdmdediipeediisrsqfpeswlwtiee 782
Qy    722 PRCIKAFTECCVVASQLRA-NISHKDMQLGRLHMK-TLLPVSKPEIRSYFPESWLWEVH- 778

Db    783 lkeperngistktmniflkdsittweilavslsdkkgicvadpfevtvmqdffidlrlpy 842
Qy    779 L-VP-RR----KQLQFALPDSLTTWEIQGIGISNT-GICVADTVKAKVFKDVFLEMNIPY 831

Db    843 svvrneqveiravlynyrea--q-slkvr-vellhnpafcslatakkrhtqtv--tigpk 896
Qy    832 SVVRGEQIQLKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTKSSKCVRQKVEGS 891

Db    897 ssvavpyvlvplkiglqevevkaavynyfisdgvkktlkvvpegmrvnktvairtlnpeq 956
Qy    892 SSHLVTFTVLPLEIGLHNINF--SLETWFGKEILVKTLRVVPEGVKRESYSGV-TLDPRG 948

Db    957 lgqggv-qreeip-aadlsdqvpdtdsetkil-lqgtpvaqmaedavdaerlkhliitps 1013
Qy    949 I-YGTISRRKEFPYRIPL-DLVPKTEI-KRILSVKGLLVGEILSAVLSQEGINILTHLPK 1005

Db   1014 gcgeqnmigmtptviavhyldqteqwekf-g--lekrqealnlinrgytqqlafkqpnwa 1070
Qy   1006 GSAEAELMSVVPVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEGMLSIMSYRNADYS 1065

Db   1071 yaafknrasstwltayvvkvfslaanligidsevlcgavkwlilekqkpdgvfqedgpvi 1130
Qy   1066 YSVWKGGSASTWLTAFALRVLGQVNKYVEQNQNSICNSLLWLVENYQLDNGSFKENSQYQ 1125

Db   1131 hqemiggvrt-aqeadvsltafvlialqeakdicraqvnnleaninkagdyiesryadvr 1189
Qy   1126 PIKLQGTLPVEARENSLYLTAFTVIGIRKAFDIC-PLVK-IDTALIKADNFLLENTLPAQ 1183

Db   1190 rpytlaiagyalaller----lng--atlqk-fl-naateknr-weea-rqk---l-ys- 1234
Qy   1184 STFTLAISAYALSLGDKTHPQFRSIVSALKREALVKGNPPIYRFWKDNLQHKDSSVPNTG 1243

Db   1235 ----veatsyallalllllkdfdavppvvrwlneqryygrgygstqatfmvfqalaqyqtd 1290
Qy   1244 TARMVETTAYALLTSLNLKDINYVNPVIKWLSEEQRYGGGFYSTQDTINAIEGLTEYSLL 1303

Db   1291 vpdhkdlnmevalqlpsrsspskfrlvweagsllrseatkqnegfkltak-gkgqgtlsv 1349
Qy   1304 VKQLR-LSMDIDVSYKHKGALHNYKMT-DKNFLGRPVEVLLNDDLIVSTGFGSGLATVHV 1361

Db   1350 vavyyaktkrkvvcknfdlrvtlkpapdtvkkpqeakstmilgictry-lg-dqdat-ms 1406
Qy   1362 TTVVHKTSTSEEVC-SFYLKIDTQDIEASHYRGYGNSDYKRIVACASYKPSREESSSGSS 1420

Db   1407 --ildismmtgfipdtddlkllatgvdryiskyemnkdfskntliiyldkvshseeecls 1464
Qy   1421 HAVMDISLPTGISANEEDLKALVEGVDQLFTDYQI-KD-G-H-VILQLNSIPSSDFLCVR 1476

Db   1465 fkihqffnvgliqpgsvkvysyynldetctqfyhpekedgmlnklchkdlcrcaeencf- 1523
Qy   1477 FRIFELFEVGFLSPATFTVYEYHRPDKQCTMFY-S-TSNIKIQKVCEGAACKCVEADCGQ 1534

Db   1524 iq--lpekitlderlekacepgvdyvyktkllkmelsddfdeyimtieqviksgsdevqa 1581
Qy   1535 MQEELDLTISAETRKQTACKPEIAYAYKVSITSITVENVFVKYKATLLDIYKTGEAVAEK 1594
```

```
Db   1582 gkerrfishikcrdalhlkegkhylmwglss-dl-wgerpnmsyiigkdtwveawpeaee 1639
             |  ||  :  | :|  |  |::||: |  : ::          | :   ||:| ||
Qy   1595 DSEITFIKKVTCTNA-ELVKGRQYLIMGKEALQIKYNFSFRYIYPLDSLTWIEYWPRDTT 1653

Db   1640 cqdeenqqqcqdlgtftenmvvfgc 1664
             |     |    :|  |:|:: : ||
Qy   1654 C-SS-CQAFLANLDEFAEDIFLNGC 1676


RESULT    9
ENTRY           C3RT       #type complete
TITLE           complement C3 precursor - rat
ALTERNATE_NAMES 37K phospholipase A2 inhibitory protein
CONTAINS        alternative-complement-pathway C3/C5 convertase (EC
                  3.4.21.47) C3b subunit; C3a anaphylatoxin
ORGANISM        #formal_name Rattus norvegicus #common_name Norway rat
DATE            04-Dec-1992 #sequence_revision 07-Oct-1994 #text_change
                  16-Feb-1997
ACCESSIONS      S15764; A54562; A01260; B35979; A35979; PN0567; PN0566;
                  A32281; S08692
REFERENCE       S15764
   #authors     Misumi, Y.; Sohda, M.; Ikehara, Y.
   #journal     Nucleic Acids Res. (1990) 18:2178
   #title       Nucleotide and deduced amino acid sequence of rat complement
                  C3.
   #cross-references MUID:90245672
   #accession   S15764
      ##molecule_type mRNA
      ##residues      1-1663 ##label MIS
      ##cross-references EMBL:X52477
REFERENCE       A54562
   #authors     Sundstrom, S.A.; Komm, B.S.; Ponce-de-Leon, H.; Yi, Z.;
                  Teuscher, C.; Lyttle, C.R.
   #journal     J. Biol. Chem. (1989) 264:16941-16947
   #title       Estrogen regulation of tissue-specific expression of
                  complement C3.
   #accession   A54562
      ##status        translation not shown
      ##molecule_type mRNA
      ##residues      'P',1316-1595 ##label SUN
      ##cross-references GB:M29866
REFERENCE       A01260
   #authors     Jacobs, J.W.; Rubin, J.S.; Hugli, T.E.; Bogardt, R.A.; Mariz,
                  I.K.; Daniels, J.S.; Daughaday, W.H.; Bradshaw, R.A.
   #journal     Biochemistry (1978) 17:5031-5038
   #title       Purification, characterization, and amino acid sequence of
                  rat anaphylatoxin (C3a).
   #cross-references MUID:79062262
   #accession   A01260
      ##molecule_type protein
      ##residues      671-703,'K',705-720,'KL',723-748 ##label JAC
      ##note          three disulfide bonds are present
REFERENCE       A35979
   #authors     Suwa, Y.; Kudo, I.; Imaizumi, A.; Okada, M.; Kamimura, T.;
                  Suzuki, Y.; Chang, H.W.; Hara, S.; Inoue, K.
   #journal     Proc. Natl. Acad. Sci. U.S.A. (1990) 87:2395-2399
   #title       Proteinaceous inhibitors of phospholipase A-2 purified from
                  inflammatory sites in rats.
   #cross-references MUID:90207203
   #accession   B35979
      ##status        preliminary
      ##molecule_type protein
      ##residues      'X',998-1005 ##label SUW
   #accession   A35979
```



```
      ##molecule_type protein
      ##residues      'X',961-962,'P',964-969 ##label SU2
REFERENCE       PN0566
   #authors     Nakagawa, H.; Komorita, N.
   #journal     Biochem. Biophys. Res. Commun. (1993) 194:1181-1187
   #title       Complement component C3-derived neutrophil chemotactic
                  factors purified from exudate of rat carrageenin-induced
                  inflammation.
   #accession   PN0567
      ##molecule_type protein
      ##residues      568-592 ##label NAK
      ##note          amino end of a C3-derived peptide designated exudate
                        neutrophil chemotactic factor 2, or C3-beta-c
   #accession   PN0566
      ##molecule_type protein
      ##residues      671-687 ##label NA2
      ##note          amino end of peptide designated neutrophil chemotactic
                        factor 1 and probably identical to C3a anaphylatoxin
REFERENCE       A32281
   #authors     Kuivanen, P.C.; Capulong, R.B.; Harkins, R.N.; DeSombre, E.R.
   #journal     Biochem. Biophys. Res. Commun. (1989) 158:898-905
   #title       The estrogen-responsive 110K and 74K rat uterine secretory
                  proteins are structurally related to complement component
                  C3.
   #cross-references MUID:89149812
   #accession   A32281
      ##molecule_type protein
      ##residues      25-41 ##label KUI
      ##experimental_source 17beta-estradiol-stimulated uterus of immature rat
      ##note          the authors treat this 74K uterine secretory protein,
                        identical as far as sequenced to complement C3, as a
                        distinct, homologous protein rather than as an
                        alternatively processed form or fragment
COMMENT    Complement C3 contains two chains, formed by removal of four
             residues and linked by a disulfide bond. Its activation by a C3
             convertase, which is the central reaction in both classical and
             alternative complement pathways, releases the C3a anaphylatoxin
             from the amino end of the alpha chain and generates C3b, which
             associates with the Bb fragment of complement factor B to form
             the alternative-complement-pathway C3/C5 convertase.
COMMENT    C3a anaphylatoxin is a vasoactive peptide and a mediator of
             inflammation.
COMMENT    C3b, with its highly reactive thiol group, binds to the surface of
             foreign particles and facilitates phagocytosis. It binds to
             complement C5 and renders it susceptible to proteolysis by the
             classical-complement-pathway C3/C5 convertase. The activity of
             C3b is regulated by proteolytic cleavage involving factors H and
             I. Its degradation products can also be biologically active.
COMMENT    The major site of synthesis of this plasma protein is the liver.
CLASSIFICATION  #superfamily alpha-2-macroglobulin
KEYWORDS        acute phase; chemotaxis; complement alternate pathway;
                  complement pathway; glycoprotein; hydrolase; immune
                  response; inflammation; liver; plasma; serine proteinase;
                  thiolester bond
FEATURE
   1-24             #domain signal sequence #status predicted #label SIG\
   25-666           #product complement C3 and C3b beta chain #status
                      predicted #label C3BB\
   25-666,671-1663  #product complement C3 #status predicted #label CC3\
   25-666,749-1663  #product complement C3b #status predicted #label C3B\
   671-1663         #product complement C3 alpha chain #status predicted
                      #label CC3A\
   671-748          #product C3a anaphylatoxin #status experimental #label
                      C3T\
```

```
    749-1663           #product complement C3b alpha' chain #status predicted
                         #label C3BA\
    946-1303           #product C3dk fragment #status predicted #label CDK\
    1002-1303          #product C3d fragment #status predicted #label C3D\
    1424-1457          #region properdin binding\
    558-816,626-661,
    693-720,694-727,
    707-728,873-1513,
    1101-1158,
    1358-1489,
    1389-1458,
    1506-1511,
    1518-1590,
    1537-1661          #disulfide_bonds #status predicted\
    748-749            #cleavage_site Arg-Ser (C3 convertase) #status
                         predicted\
    939,1617           #binding_site carbohydrate (Asn) (covalent) #status
                         predicted\
    1010-1013          #cross-link thiolester (Cys-Gln) #status predicted\
    1303-1304          #cleavage_site Arg-Ser (complement factor I) #status
                         predicted\
    1320-1321        • #cleavage_site Arg-Ser (complement factor I) #status
                         predicted
SUMMARY          #length 1663  #molecular-weight 186459  #checksum 3009

  Query Match            18.7%;  Score 2255;  DB 2;  Length 1663;
  Best Local Similarity  28.7%;  Pred. No. 0.00e+00;
  Matches  478;  Conservative  422;  Mismatches 665;  Indels 102;  Gaps  76;

Db     28 ysiitpnvlrleseetfileahdaqgdvpvtvtvqdfl-kkqvltsektvltgatghlnr 86
           | | :|:::|: : | ::::           |:::  :  ||   :|    |::     |
Qy     23 YVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSYPDKKFSYSSGHVHLSSENKFQNS 82

Db     87 vfikipaskefnadkgh-kyvtvvanfgatvvekavlvsfqsgylfiqtdktiytpgstv 145
           :: | : |:: :        || :          | : ::::|:||:|:|| :||| :|
Qy     83 AILTIQP-KQLPGGQNPVSYVYLEVVSKHFSKSKRMPITYDNGFLFIHTDKPVYTPDQSV 141

Db    146 fyriftvdnnllpvgktvvivietpdgvpikrdilsshnqygilpls-wnipelvnmgqw 204
           |::::::::| |  :  |: :  |:|  :  |::    :: ||:::: :||     | |
Qy    142 KVRVYSLNDDLKPAKRETVLTFIDPEGSEV--DMVEEIDHIGIISFPDFKIPSNPRYGMW 199

Db    205 kirafyehapkqtfsaefevkeyvlpsfevlveptekfy-ylhgpkglevsitarflygk 263
           |:| |       | :| ||||||||| | | :||   :|   | :    | :|::| ||::| |
Qy    200 TIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEPEYNFIGYKNF-KNFEITIKARYFYNK 258

Db    264 nv-dgtafvifgv-qd--edkkislalsltrvliedgsgeavlsrkvlmdgvrpsspeal 319
           | ::  :: ||: :|  :| |  :  ::    :: :| ::  :       :   :  | | |
Qy    259 VVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLINGIAQVTFDSETAVKELSYYSLEDL 318

Db    320 vgkslyvsvtvilhsgsdmveaersgipivtspyqihftktpkffkpampfdlmvfvtnp 379
            | ||::||||   :|:    |||:||  | ||| :::  || |:||::|: ::| | ::
Qy    319 NNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLNLVATPLFLKPGIPYPIKVQVKDS 378

Db    380 -d---gs-par-rvp-v-vtqg-sd---aqaltq-ddgvaklsvntpnnrqpltitvstk 426
           |    |: |          :  | |   ||    : ::|: |||||  : :| |:    | : | |
Qy    379 LDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDDGVASFVLNLPSGVTVLEFNVKTD 438

Db    427 kegipdarqatrtmqaqpystmhnsnnylhlsvsrvelkpgdnlnvnfhlrtdaqqeaki 486
           :|:   ||     :| :|::|:  |  |:  : ::  |  |::||: : :     :    ||
Qy    439 APDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKALLVGEHLNI-I-VTPKSPYIDKI 496

Db    487 ryytylvmnkgkllkagrqvrepqqdlvvlslpitpefipsfrlvayytliganqqrevv 546
            | ||:::|||::  | : : :          :::|:| : :|| ||: || : |      |:|
Qy    497 THYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQNMVPSSRLLVYYIVTGEQTA-ELV 555
```



```
Db    547 adsvwvdvkdscvgtlvvkgdprdnrqpapghqttlriegnqgarvglvavdkgvfvlnk 606
           :||||::: : |    | |    | |      :||:  :| :      ::|:| ||| :|: : :
Qy    556 SDSVWLNIEEKCGNQLQVHLSP-DADAYSPGQTVSLNMATGMDSWVALAAVDSAVYGVQR 614

Db    607 knkltqskiwdvvekadigctpgsgknyagvfmdagltfktnqglqtdqredpecakpaa 666
            |     ::  : :||:|:|| :|:| | | || |||| ||   :  | :|  | |
Qy    615 GAKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLAGLTFLTNANADDSQENDEPC-KEIL 673

Db    667 rrrrsvqlmerrmdkaqqytdkglrkccedgmrdipmpyscqrrarlitqgesclkafmd 726
           | ||::| : : : |: |     ::||| ||   :     :|::|| |: | |:||| :
Qy    674 RPRRTLQ-K-KIEEIAAKYKHSVVKKCCYDGAC-VNNDETCEQRAARISLGPRCIKAFTE 730

Db    727 ccnyitklreqhrrdhvlglarsdvdediipeediisrshfpeswlwtieelkepekngi 786
           ||   : ||    :    : |:|  :    ::|        || ||||||| :  |  | :
Qy    731 CCVVASQLRANISHKD-MQLGRLHMK-TLLPVSKPEIRSYFPESWLWEVH-L-VP-RR-- 783

Db    787 stkvmniflkdsittweilavslsdkkgicvadpyeitvmqdffidlrlpysvvrneqve 846
           | : : | ||:||||| ::::|:   |||||        |  | |::: :|||||| ||::
Qy    784 --KQLQFALPDSLTTWEIQGIGISNT-GICVADTVKAKVFKDVFLEMNIPYSVVRGEQIQ 840

Db    847 iravlfnyreqek---lkvr-vellhnpafcsmatakkryyqti-e-ippkssvavpyvi 900
           ::: ::|||    :     :|:   || :    :      :          : : :   || | : :
Qy    841 LKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTKSSKCVRQKVEGSSSHLVTFTV 900

Db    901 vplkiglqevevkaavfnhfisdgvkkilkvvpegmrvnktvavrtldpehlnqggvqre 960
           :|| |||::::          |   : : | |:|||||:: :   :| ||||  :   | : |
Qy    901 LPLEIGLHNINFSLETW--FGKEILVKTLRVVPEGVKRESYSGV-TLDPRGIY-GTISRR 956

Db    961 dvnaadls-dqvpdtdsetril-lqgtpvaqmaedavdgerlkhlivtpsgcgeqnmigm 1018
             :  :: | || |:    ||| : |  |:::     :    | :: |    | | :| :::::
Qy    957 KEFPYRIPLDLVPKTEI-KRILSVKGLLVGEILSAVLSQEGINILTHLPKGSAEAELMSV 1015

Db   1019 tptviavhyldqteqwekf-g--lekrqealelikkgytqqlafkqpisayaafnnrpps 1075
           |       |||:  ::|:  | :   |  :|      :| |     :::: :   :|:   :  ::|
Qy   1016 VPVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEGMLSIMSYRNADYSYSVWKGGSAS 1075

Db   1076 twltamwsrsfslaanliaidsqvlcgavkwlilekqkpdgvfqedgpvihqemiggfr- 1134
           |||||    | ::    : :  :     :| :: ||: : |   :| | |::        : | :
Qy   1076 TWLTAFALRVLGQVNKYVEQNQNSICNSLLWLVENYQLDNGSFKENSQYQPIKLQGTLPV 1135

Db   1135 ntkeadvsltafvlialqeardicegqvnslpgsinkageyleasylnlqrpytvaiagy 1194
           :::|  : |||| :|::: ||| |||     |: :   :: || ::|  : |    :|:||:|
Qy   1136 EARENSLYLTAFTVIGIRKAFDICP-LVK-IDTALIKADNFLLENTLPAQSTFTLAISAY 1193

Db   1195 alalmnkleepy--ltkflnt-a--k-drn--r-weepgqq----l-yn-----veatsy 1235
           ||:| :|  .  :   :   |:   |   |  :      | |  :   |:     :          ||:|:|
Qy   1194 ALSLGDKTHPQFRSIVSALKREALVKGNPPIYRFWKDNLQHKDSSVPNTGTARMVETTAY 1253

Db   1236 allallllkdfdsvppvvrwlnderyyggygygstqatfmvfqalaqyradvpdhkdlnmd 1295
           |||: | |||:: | |||::|||::|: ||||: ||| |:  :::|::|    | : : |:||
Qy   1254 ALLTSLNLKDINYVNPVIKWLSEEQRYGGGFYSTQDTINAIEGLTEYSLLVKQLR-LSMD 1312

Db   1296 vslhlpsrssptvfrllwesgsllrseetkqnegfsltak-gkqqgtlsvvtvyhakvkg 1354
           : :    :::    ::: :     | |: |     |: : :::  | | :|: | || |    :
Qy   1313 IDVSYKHKGALHNYKMT-DKNFLGRPVEVLLNDDLIVSTGFGSGLATVHVTTVVHKTSTS 1371

Db   1355 kttckkfdlrvtikpapetakkpqdakssmildictry-lg-dvdat-ms--ildismmt 1409
             |  | |::                     :  :  |: |  :  :  ::  |  ::|||: |
Qy   1372 EEVCS-FYLKIDTQDIEASHYRGYGNSDYKRIVACASYKPSREESSSGSSHAVMDISLPT 1430

Db   1410 gfipdtndlellssgvdryiskyemdkafsnkntliiylekishseedclsfkvhqffnv 1469
           |: :: :||· |  |||: :: |::  |   ::   :|: |: |: |:  |: |:: ::|:|
Qy   1431 GISANEEDLKALVEGVDQLFTDYQI-KD-GH---VILQLNSIPSSDFLCVRFRIFELFEV 1485
```

```
Db   1470 gliqpgsvkvysyynleesctrfyhpekddgmlsklchnemcrcaeencf-mhqsqd-qv 1527
          |:: |:: || |    :  || || :   : :: |:|    |:| | :| |:: |  :
Qy   1486 GFLSPATFTVYEYHRPDKQCTMFY-S-TSNIKIQKVCEGAACKCVEADCGQMQEELDLTI 1543

Db   1528 slnerldkacepgvdyvykt kltt ielsddfdeyimtieqviksgsdevqagqerrfish 1587
           | : | : || | : | ||  :|:| : : |  |  |: :: |:|    :   |  ||
Qy   1544 SAETRKQTACKPEIAYAYKVSITSITVENVFVKYKATLLDIYKTGEAVAEKDSEITFIKK 1603

Db   1588 vkcrnalklqkgkqylmwglss-dl-wgekpntsyiigkdtwvehwp 1632
           | | ||  | ||:|||: |   : ::          | :   ||:| ||
Qy   1604 VTCTNA-ELVKGRQYLIMGKEALQIKYNFSFRYIYPLDSLTWIEYWP 1649
```

```
RESULT   10
ENTRY           I51018      #type complete
TITLE           cobra venom factor precursor - Naja naja
ORGANISM        #formal_name Naja naja
DATE            13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
                  16-Feb-1997
ACCESSIONS      I51018
REFERENCE       I51018
   #authors     Fritzinger, D.C.; Bredehorst, R.; Vogel, C.W.
   #journal     Proc. Natl. Acad. Sci. U.S.A. (1994) 91:12775-12779
   #title       Molecular cloning and derived primary structure of cobra
                  venom factor.
   #cross-references MUID:95108041
   #accession   I51018
      ##status        preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues      1-1642 ##label FRI
      ##cross-references EMBL:U09969; NID:g881914; CDS_PID:g881915
CLASSIFICATION  #superfamily alpha-2-macroglobulin
SUMMARY         #length 1642  #molecular-weight 184516  #checksum 1347

  Query Match            18.4%;  Score 2214;  DB 6;  Length 1642;
  Best Local Similarity  29.3%;  Pred. No. 0.00e+00;
  Matches  502;  Conservative  438;  Mismatches  642;  Indels  132;  Gaps  103;
```

```
Db      4 malylvaalligfpg-s-shgalytlitpavlrtdteeqilveahgdstpkqldifvhdf 61
          |:| :  :|| | | : ::   | : :| ::| : | |::: | : : : | : :  :
Qy      1 MGLLGILCFLI-FLGKTWGQEQTYVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSY 59

Db     62 prkqktlfqtrvdmnpaggmlvtptieipakevstdsrqnqyvvvqvtgpqvrlekvvll 121
          | |  :    :| :::      : : : :|:::        || ::| : :    | : :
Qy     60 PDKKFSYSSGHVHLSSENKFQNSAILTIQPKQLPGGQNPVSYVYLEVVSKHFSKSKRMPI 119

Db    122 syqssflfiqtdkgiytpgspvlyrvfsmdhntskmnktvivefqtpegilvs-snsvd- 179
          :|:::||||:||| :||| :|  ||:|:: :    ::  :: |  |||  |    : :|
Qy    120 TYDNGFLFIHTDKPVYTPDQSVKVRVYSLNDDLKPAKRETVLTFIDPEGSEVDMVEEIDH 179

Db    180 lnff-wp-ynlpdlvslgtwrivaky-ehspenytayfdvrkyvlpsfevrlqpsekff- 235
          : ::  | :::|       | | | ||| |  :    ||||:|: |||| | | ::| :|:
Qy    180 IGIISFPDFKIPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEPEYNFIG 239

Db    236 yidgnenfhvsitarylygeev-egvafvlfgvk--i-ddakksipdsltripiidgdgk 291
          | : : || ::| |||:|   | |:  :: ||::  : || |   :  ::     :|:| :
Qy    240 YKNFK-NFEITIKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLINGIAQ 298

Db    292 atlkrdt-frs-rfpnlnelvghtlyasvtvmtesgsdmvvteqsgihivaspyqihftk 349
          |:  :| :   : :|::|     || :|||:  :|:    :| :||  | ||| :::
Qy    299 VTFDSETAVKELSYYSLEDLNNKYLYIAVTVIESTGGTSEEAEIPGIKYVLSPYKLNLVA 358

Db    350 tpkyfkpgmpyeltvyvtnp-d---gs-paa-hvp-v-vs-ea--fh-smgtt-lsdgta 396
```



```
          || ::|||:|| : | | :: |   |: |    :    : |: |:  :  |:: | : || |
Qy    359 TPLFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDDGVA 418

Db    397 klilnipInaqslpitvrtnhgdlprerqatksmtaiayqtqggsgnylhvaitsteikp 456
          ::||:|     | : |:|:  ||| | ||   |||| : : |  |:  :        :
Qy    419 SFVLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKALLV 478

Db    457 gdnlpvnfnvkgnanslkqikyftylilnkgkifkvgrqprrdgqnlvtmnlhitpdlip 516
          |::| : : |  ::  :   |  : ||||:||||:  | : :     :  ::|: :| :::|
Qy    479 GEHLNI-I-VTPKSPYIDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQNMVP 536

Db    517 sfrfvayyqv-gnn--eivadsvwvdvkdtcmgtl-v-vkgdnliqmpgaamkiklegdp 571
          | |:: || | |:    |:|:||||::: : |    | | :   |      || :: :::
Qy    537 SSRLLVYYIVTGEQTAELVSDSVWLNIEEKCGNQLQVHLSPDADAYSPGQTVSLNMATGM 596

Db    572 garvglvavdkavyvlndkykisqakiwdtieksdfgctagsgqnnlgvfedaglaltts 631
          ::|:| ||| ||| :     | :   :: : :|||:|| ||:| ||   ||  |||:: |:
Qy    597 DSWVALAAVDSAVYGVQRGAKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLAGLTFLTN 656

Db    632 tnlntkqrsaakcpqpanrrrrssvllldsnaskaaefqdqdlrkccedvmhenpmgytc 691
          :| :  | :'  | :   | ||:   |          || :    ::||| |     |    ||
Qy    657 ANADDSQENDEPCKE-ILRPRRT---LQKKIEEIAAKYKHSVVKKCCYDGACVNNDE-TC 711

Db    692 ekrakyiqegdackaaflecc ryikgvrdenqreselflarddnedgfiadsdiisrsdf 751
          | ||  |  |  | || |||      :| |    :: |:|  :   ::: |     || |
Qy    712 EQRAARISLGPRCIKAFTECCVVASQLRA-NISHKDMQLGRL-HMKTLLPVSKPEIRSYF 769

Db    752 pkswlwltkdlteepnsqgissktmsfylrdsittwvvlavsftptkgicvaepyeirvm 811
          | ||||    |   |  :    | : | | ||:|||  : ::::: | |||||:     :|
Qy    770 PESWLWEV-HLV--PR-R----KQLQFALPDSLTTWEIQGIGISNT-GICVADTVKAKVF 820

Db    812 kvffidlqmpysvvkneqveirailhnyvnedi-y-vr---velly---npafcsastkg 863
          |  |::: :|||||: ||::::: : ||    : : |:    || :     :| :   :||:
Qy    821 KDVFLEMNIPYSVVRGEQIQLKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTKS 880

Db    864 qryrqqfpikalssravpfvivpleqglhdveikasvqealw-sdgv-rkklkvvpegvq 921
          :  :|   :  : ||: | |  : :|||  ||||::::  |: |: |  : :   | |:||||||
Qy    881 SKCVRQ-KVEGSSSHLVTFTVLPLEIGLHNINF--SL-ET-WFGKEILVKTLRVVPEGVK 935

Db    922 ksivtivkldprak-gvggtqlevikarklddrvpdteietkiiiqgdpvaqiiensidg 980
           :  : ||||:  : : | : |  | || |||   : : |  |::|:   :
Qy    936 RESYSGVTLDPRGIYGTISRRKE-FPYRIPLDLVPKTEIKRILSVKGLLVGEILSAVLSQ 994

Db    981 sklnhliitpsgcgeqnmirmaapviaty-yldttteqwetlginrrte-a-vnqivt-gy 1036
          :| |  | | :| ::: : ||:  :  ||:| ::|: : :  :     |   ::  :  |
Qy    995 EGINILTHLPKGSAEAELMSVV-PVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEGM 1053

Db   1037 aqqmvykkadhsyaaftnrassswltayvvkvfamaakmvagisheiicggvrwlilnrq 1096
           | |::|| ||:     ::|:||||: ::|::  | |   ::: || :: ||: | |
Qy   1054 LSIMSYRNADYSYSVWKGGSASTWLTAFALRVLGQVNKYVEQ-NQNSICNSLLWLVENYQ 1112

Db   1097 qpdgafkenapvlsgtmqgg--iqgaeeevyltafilvalleskticndyvnsldssikk 1154
           :|:||||:  :  :||   ::: |: :||||| :::: :  ||   |: :|::: |
Qy   1113 LDNGSFKENSQYQPIKLQGTLPVEARENSLYLTAFTVIGIRKAFDICP-LVK-IDTALIK 1170

Db   1155 atnyllkkyeklqrpyttaltayalaaadqlndd-r--v--lm-aa-stgrd---h-w-e 1202
          | |:|| :    | :| |::||||: :|  : : |  | |:  |   |      : | :
Qy   1171 ADNFLLENTLPAQSTFTLAISAYALSLGDKTHPQFRSIVSALKREALVKGNPPIYRFWKD 1230

Db   1203 --ey-na---ht---hniegtsyallallkmkkfdqtgpivrwltdqnfygetygqtqat 1253
           :  ::   :|   : :| |:||||: |::| ::   |:::||:::  ||  :  || |
Qy   1231 NLQHKDSSVPNTGTARMVETTAYALLTSLNLKDINYVNPVIKWLSEEQRYGGGFYSTQDT 1290

Db   1254 vmafqalaeyeiqmpthkdlnlditielpdrevpiryrinyenallartvetklnqditv 1313
```

```
                 : |:::|:|| : :  :  |::|| :     :      |::  :: :|:| ||  ||:|: |
Qy    1291 INAIEGLTEYSL-LVKQLRLSMDIDVSYKHKGALHNYKMT-DKNFLGRPVEVLLNDDLIV 1348

Db    1314 tas-gdgkatmtiltfynaqlqekanvcnkfhlnvsvenihlna-mg-akgalmlki-ct 1369
           ::: | | ||: : |  :           :||: | |::  ::|  :    | ::: :  : |:
Qy    1349 STGFGSGLATVHVTTVVH-KTSTSEEVCS-FYLKIDTQDIEASHYRGYGNSDYKRIVACA 1406

Db    1370 ry-lg--evds-tm-tiidismltgflpdaedltrlskgvdryisryevdnnmaqkvavi 1424
             |  :  |  | :  :::|||: ||: :: |||  |  |||: :: |:: :: ::   ||
Qy    1407 SYKPSREESSSGSSHAVMDISLPTGISANEEDLKALVEGVDQLFTDYQI-KD-GH---VI 1461

Db    1425 iylnkvshsedeclhfkilkhfevgfiqpgsvkvysyynldekctkfyhpdkgtgllnki 1484
           : || :: |:  |::|:|:  |||||: |::  || |   |  ||:||   :   :  |:
Qy    1462 LQLNSIPSSDFLCVRFRIFELFEVGFLSPATFTVYEYHRPDKQCTMFYSTSN-IKI-QKV 1519

Db    1485 cignvcrcagetcsslnhq-e-ridvplqiekacetnvdyvyktkllrieeqdgndiyvm 1542
           | |  |:|    |: :  : :  |     : : ||  :: | ||   :   |  ::     |
Qy    1520 CEGAACKCVEADCGQMQEELDLTISAETRKQTACKPEIAYAYKVSITSITVENVFVKYKA 1579

Db    1543 dvlevikqgtdenprakthqyisqrkcqealnlkvnddyliwgsrsdllptk-d-kisyi 1600
            :|:: | |      :      :|     | :| :|     :||| | : : |  | :  : ||
Qy    1580 TLLDIYKTGEAVAEKDSEITFIKKVTCTNA-ELVKGRQYLIMG-K-EALQIKYNFSFRYI 1636

Db    1601 it-kn-twierwphedecqe-eef-qklcddfaq 1630
               : |||| ||::  |   : |  :| |:||:
Qy    1637 YPLDSLTWIEYWPRDTTCSSCQAFLANL-DEFAE 1669
```

```
RESULT   11
ENTRY            S21045      #type fragment
TITLE            complement protein homolog - inshore hagfish (fragment)
ORGANISM         #formal_name Eptatretus burgeri #common_name inshore hagfish
DATE             13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
                   13-Jan-1995
ACCESSIONS       S21045
REFERENCE        S21045
   #authors      Ishiguro, H.; Kobayashi, K.; Suzuki, M.; Titani, K.;
                   Tomonaga, S.; Kurosawa, Y.
   #journal      EMBO J. (1992) 11:829-837
   #title        Isolation of a hagfish gene that encodes a complement
                   component.
   #cross-references MUID:92192016
   #accession    S21045
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-1620 ##label ISH
      ##cross-references EMBL:Z11595
SUMMARY          #length 1620  #checksum 3983

  Query Match           17.5%;  Score 2106;  DB 15;  Length 1620;
  Best Local Similarity  29.6%;  Pred. No. 0.00e+00;
  Matches  466;  Conservative  416;  Mismatches  588;  Indels  104;  Gaps   88;

Db      93 ipltsrrglvfaqtdqpiytpnndvnirlfpvtrqlnpilsslvvdimnpdgvvvdriek 152
           :|:|  |::| :|| |:|||:  |::|::::  :|:|        |: :::|:|  || :|
Qy     117 MPITYDNGFLFIHTDKPVYTPDQSVKVRVYSLNDDLKPAKRETVLTFIDPEGSEVDMVEE 176

Db     153 nafevekvmelrpfhvpaitslgdwkivswmkdkpqfnyt-sg-fkveeyvlptfdvsit 210
             :  :: :   | :|:       | | |  :  | | :|: | :: | | ||||| | |||
Qy     177 ID-HIG-IISFPDFKIPSNPRYGMWTIKA--KYKEDFSTTGTAYFEVKEYVLPHFSVSIE 232

Db     211 seqpylhvy-d-kaftihikamhiygkpvm-grayvrygvkhqsk---rtllstssalar 264
           :|  ::  | : | | | | ||| :| | |  :  |: :|:: : |     : :: |:   :
Qy     233 PEYNFI-GYKNFKNFEITIKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTM 291
```



```
Db     265 feqgeamhtl-rqkhile-qypdpklllgqslyveasvissdageiensilddipivasp 322
           :  | |  |:  :  : |  |       |     ||:  :|| | :|  |:: :   |  | ||
Qy     292 LINGIAQVTFDSETAVKELSYYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSP 351

Db     323 ysikskwtvpffkpgvpyiykvlvlnpdgspasgvp-i-kv-sfsfd-ssgnwitqkrkt 378
           | ::    |  |:|||:||  || |  ::       :||| | :  ::  :  :::     |  |
Qy     352 YKLNLVATPLFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVT 411

Db     379 -mdngiamqtintarnskklnikvqtederleqsqqaeasftiasysspsgsfihln-a- 435
            :|:|:|    :| :     |:::| |:   | :  ||  ::    :||| | |:: :: :
Qy     412 RVDDGVASFVLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTD 471

Db     436 hrevkspgehivfdvfiksaakdhvlhfnylmisnqkihnflqegrkgdtt--svslllt 493
           ::       |||: : |   ||:   | :  |:|||::|:|||  :|     : :|::  |::: :|
Qy     472 NHKALLVGEHLNIIVTPKSPYIDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVT 531

Db     494 pelvpqfrlvaffilp---sgelvadsiiidvkdschaklsldvaggkrlfsprdnvnfd 550
           ::|| ||: ::|:     ::|||:||: ::: : |     | : ::       :|| : |:::
Qy     532 QNMVPSSRLLVYYIVTGEQTAELVSDSVWLNIEEKCGNQLQVHLSPDADAYSPGQTVSLN 591

Db     551 lsgesdswvavgvvdkaayvldkknkltankvykameasdlgcsvgsgktgplvfrdagl 610
           ::     |||||::  || |  | :::   |        ::|:  :| |||||: |:|   : ||: |||
Qy     592 MATGMDSWVALAAVDSAVYGVQRGAKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLAGL 651

Db     611 aimakeisgmddvkdpgcpnghtrrkrelvleiaiekasty-paelrkccrdaaiesplr 669
           :::::   :     : :|   | :    | :| |     |   | |: |  : ::||| |:|  :
Qy     652 TFLTNANADDSQENDEPC-KEILRPRRTLQKKIE-EIAAKYKHSVVKKCCYDGACVNN-D 708

Db     670 lsceertkhihdegegcqetflecckhveeelliameeededlgrsqgedfmiqesqvvi 729
            :||:|: :|    |   |  :| ||| |   :|    :    | :||| :       :: |    |
Qy     709 ETCEQRAARI-SLGPRCIKAFTECCV-VASQLRANISHKDMQLGRLHMKT-LLPVSKPEI 765

Db     730 rshfpesfmweiiklsrsaengksritkkmpdsittwdiqavevsqskglcvgpsleltv 789
           || |||| :||: ::|   :   :::  :|||:|||:||::  :| : |:||: ::   |
Qy     766 RSYFPESWLWEVHLVPRR--K-QLQFA--LPDSLTTWEIQGIGISNT-GICVADTVKAKV 819

Db     790 fkqfflkvhtpyalkqyeqvelrvviynymnqdvkgeiqvkcgdgict-dae--qneplk 846
           ||: || :: ||:: : ||::|:  :|||      :    :  :   :|||| ::    ::: |
Qy     820 FKDVFLEMNIPYSVVRGEQIQLKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTK 879

Db     847 srfav-ek-nsats-f-sfmvvplsssdssvs-vlarvfgsdvhdavekdlrvmpegnye 901
           |     | :|  ::::| : :| |:||   :  :::   |     || ::    | |  |||:|||
Qy     880 SSKCVRQKVEGSSSHLVTFTVLPLEIGLHNINFSLETWFGKEI--LV-KTLRVVPEGVKR 936

Db     902 emsrswsvqpr-rhgg-qqvivvdnetpqnvvpgtemsaflsaqgnlvaetiqntlkgsk 959
           |   : :::||    |   :         | ::|| ||:  :||  | ||:| :    |
Qy     937 ESYSGVTLDPRGIYGTISRRKEFPYRIPLDLVPKTEIKRILSVKGLLVGEILSAVLSQEG 996

Db     960 isnllrlprgcgeqnmmytsitvm-varylnrsdqwnkm-gdpqlkkrsfdf-itsgfas 1016
           |: | :||:| :| ::| : : |   | :||: :::||   :|| : |: :   :   |  |
Qy     997 INILTHLPKGSAEAELM-SVVPVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEGMLS 1055

Db    1017 qltyrkpdysyaawlhrasstwltafvakvfsqarqlvfipvseicgsvrwlmrk-qdkd 1075
           ::||::||||: |    ::|||||||  :|::|    |      :  || |: ||: : |  :
Qy    1056 IMSYRNADYSYSVWKGGSASTWLTAFALRVLGQVNKYVEQNQNSICNSLLWLVENYQLDN 1115

Db    1076 gsfle-sk-pvvhln-mmg-qvtgkvv-ltsfvfiallearesc-inevegftvvvekah 1129
           ||| | |         :  |     :  :   : :  ::||:|  |::  | : | :  : |  :: |  :: ||
Qy    1116 GSFKENSQYQPIKLQGTLPVEARENSLYLTAFTVIGIRKAFDICPLVKID--TALI-KAD 1172

Db    1130 gyltsqamngledfplaitayalslw-kv-sdgaakvt-m-h-tl-k-tsgl-q--teel 1179
           :|    :: :    | |||:||||||  |   ::   :  |: : : :| |   : : :   ::|
Qy    1173 NFLLENTLPAQSTFTLAISAYALSLGDKTHPQFRSIVSALKREALVKGNPPIYRFWKDNL 1232
```

```
Db   1180 ihw-gs--nkgkaaavestaygllaaiqheegeiaekatnwlsqsatfggyfqstqdtvm 1236

Qy   1233 QHKDSSVPNTGTARMVETTAYALLTSLNLKDINYVNPVIKWLSEEQRYGGGFYSTQDTIN 1292

Db   1237 alqaltgfescqsrmkkmdlsfkiraeengvfdkefqitndnafvqkpfkvpvhgqltvt 1296

Qy   1293 AIEGLTEY-SLLVKQLRLSMDIDVSYKHKGALHN-YKMTDKN-FLGRPVEVLLNDDLIVS 1349

Db   1297 asgtgqgiltf-vkkyrekvvikkd-ckgfsleittnldnqvkqrrrqsinpefnvyrfi 1354

Qy   1350 T-GFGSGLATVHVTTVVHKTSTSEEVC-SFYLKIDTQ-DIEASHYRGYG-NSDYK--RIV 1403

Db   1355 gcfry-l-rnqep-gm--vvmdislptgfeakkkdlddmknlvdnyivqyeirpgrvfly 1409

Qy   1404 ACASYKPSREESSSGSSHAVMDISLPTGISANEEDLKALVEGVDQLFTDYQIKDGHVILQ 1463

Db   1410 ldkvnkdekncvgfrlnqvfesnlvlpvtatvfeyyepdfrcsksyhpkmevnpdaschg 1469

Qy   1464 LNSIPSSDFLCVRFRIFELFEVGFLSPATFTVYEYHRPDKQCTMFY-STSNIKIQKVCEG 1522

Db   1470 nicnclqrhcvelkgmadedrnad-rngnacrae--yvfiigvtkvtktasyininaalk 1526

Qy   1523 AACKCVEADCGQMQEELDLTISAETRKQTACKPEIAYAYKVSITSITVENVFVKYKATLL 1582

Db   1527 tvlkkgmdqainvgarrsfvipmhcgknlnvspgdiylvmgmhnahwr-nsd-rtqyvlt 1584

Qy   1583 DIYKTG-EAVAEKDSEITFIKKVTC-TNAELVKGRQYLIMGKEALQIKYNFSFRYIYPLD 1640

Db   1585 sdtwfekfplesvc 1598

Qy   1641 SLTWIEYWPRDTTC 1654


RESULT    12
ENTRY           I51339     #type fragment
TITLE           complement component C3 - rainbow trout (fragment)
ORGANISM        #formal_name Oncorhynchus mykiss #common_name rainbow trout
DATE            13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
                  16-Feb-1997
ACCESSIONS      I51339
REFERENCE       I51339
   #authors     Lambris, J.D.; Lao, Z.; Pang, J.; Alsenz, J.
   #journal     J. Immunol. (1993) 151:6123-6134
   #title       Third component of trout complement. cDNA cloning and
                  conservation of functional sites.
   #cross-references MUID:94065166
   #accession    I51339
      ##status        preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues      1-1620 ##label LAM
      ##cross-references GB:L24433; NID:g431607; CDS_PID:g431608
CLASSIFICATION  #superfamily alpha-2-macroglobulin
SUMMARY         #length 1620  #checksum 2588

  Query Match            16.1%;  Score 1944;  DB 6;  Length 1620;
  Best Local Similarity  28.6%;  Pred. No. 0.00e+00;
  Matches  481;  Conservative  421;  Mismatches  650;  Indels  132;  Gaps   89;

Db     11 plnvkimvknhptqskelasksvvldqannfqamtql-icrgplvddpkqkqyvvlqaqf 70

Qy     49 AFDATISIKSYPDKKFSYSSGHVHLSSENKFQNSAILTIQPKQLPGGQNPVSYVYLEVVS 108

Db     71 pdrllekvvlvsfqsgyifiqtdktiytpastvhyrvfsmtpglepltreifedqevakn 130
```



```
Qy    109 KHFSKSKRMPITYDNGFLFIHTDKPVYTPDQSVKVRVYSLNDDLKPAKRETVLTFIDPEG 168

Db    131 keiavsveimtpenitifreivnpdkgvksgqfklpdivsfgtwhvvtrfqstpqktfss 190

Qy    169 SEVDM-VEEI--DHIGIIS-F--PD-------FKIPSNPRYGMWTIKAKYKEDFSTTGTA 215

Db    191 efevkeyvlpsfevsltpakaff-yvddndltvditarylygkevtgtg-yvvfgvi-t- 246

Qy    216 YFEVKEYVLPHFSVSIEPEYNFIGYKNFKNFEITIKARYFYNKVVTEADVYITFGIREDL 275

Db    247 tesekksfpaslqrveikdgkgvaclkkeh-itq-tfpkihdlvkqsifvsvsvlteggg 304

Qy    276 KDDQKEMMQTAMQNTMLINGIAQVTFDSETAVKELSYYSLEDLNNKYLYIAVTVIESTGG 335

Db    305 emveaekrgiqivtspysilfkrtpkyfkpgmpfdvsvyitnp-dn---s-pai-gve-v 357

Qy    336 FSEEAEIPGIKYVLSPYKLNLVATPLFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTI 395

Db    358 ev---tpd-h-akgvtran-gfakiplntvasatelvitvktkdpgdprqqtgggtmkal 411

Qy    396 DVNQETSDLDPSKSVTRVDDGVASFVLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAI 455

Db    412 pyrtstknflhvg-vdsn-elkigdpikidlnlgpttipn-hdltymflsrgqlvkvg-r 467

Qy    456 AYSSLSQSYLYIDWTDNHKALLVGEHLNIIVTPKSPYIDKITHYNYLILSKGKIIHFGTR 515

Db    468 fkrqgnalvtlsvpvskellpsfrivayyhv-g--aadlvadsvwvdikvscmgslkvts 524

Qy    516 EKFSDASYQSINIPVTQNMVPSSRLLVYYIVTGEQTAELVSDSVWLNIEEKCGNQLQVHL 575

Db    525 trpkasyeprrafsltitgdpgakvglvavdkgvyvlnskhrltqtkiwdtiekhdtgct 584

Qy    576 SPDADAYSPGQTVSLNMATGMDSWVALAAVDSAVYGVQRGAKKPLERVFQFLEKSDLGCG 635

Db    585 agggadnmgvfydaglvfetntakgtgirtdpscpvssrrrravtisdvitsmasky-hg 643

Qy    636 AGGGLNNANVFHLAGLTFLTNANADDSQENDEPCKEILRPRR--TLQKKIEEIAAKYKHS 693

Db    644 lakeccvdgmrdntmgytcdrraqyisdgdvcvqaflvcctemaskkieskqdalllsrs 703

Qy    694 VVKKCCYDGACVNNDE-TCEQRAARISLGPRCIKAFTECCVVASQLRANISHKDMQLGRL 752

Db    704 eedddddaymrsedivsrsqfpeswmwedtnlpecpaqnkhcestsvirnnflkdsittw 763

Qy    753 HMKTLLPV-SKPE-I--RSYFPESWLWE-VHLV--P-RRKQ------LQFA-LPDSLTTW 797

Db    764 qitaislskthgicvadpfemivlkeffidlklpysavrneqlevkailhnysedpiivr 823

Qy    798 EIQGIGISNT-GICVADTVKAKVFKDVFLEMNIPYSVVRGEQIQLKGTVYNYRTSGMQFC 856

Db    824 velmengevcssas----kkg-kyrqev--nmdpmstrvvpyviipmklglhsievkasv 876

Qy    857 VKMSAVEGICTSESPVIDHQGTKSSKCVRQKVEGSSSHLVTFTVLPLEIGLHNINF--SL 914

Db    877 knsgsndgvkrdlrvvaegvlvkketnvllnp-vklgg-eqtshipsgvprnqvpnsdad 934

Qy    915 ETWFGKEILVKTLRVVPEGVKRESYSGVTLDPRGIYGTISRRKEFPYRIPLDLVPKTEIK 974

Db    935 tlisvtageqtsvlveqaisgdslgslivqpvgcgeqnmiymtlpviat-hyldntkkwe 993

Qy    975 RILSVK-GLLVGEILSAVLSQEGINILTHLPKGSAEAELMSV-VPVFYVFHYLETGNHWN 1032

Db    994 di-g--ldkrntaikyinigyqrqlayrkedgsyaawvsrqsstwltayvvkvfamsstl 1050
```

```
            : :  |  :      | :: |     ::||: | ||: | :  :||||||: ::|::  :
Qy    1033 IFHSDPLIEKQKLKKKLKEGMLSIMSYRNADYSYSVWKGGSASTWLTAFALRVLGQVNKY 1092

Db    1051 isvqenvlctavkwlilntqqpdgifnefapvihaemtgn--vrgsdndasmtafvliam 1108
             :  :| :| :: ||: | |   :| |:| :          : |   | : :|   :||| :|::
Qy    1093 VEQNQNSICNSLLWLVENYQLDNGSFKENSQYQPIKLQGTLPVEARENSLYLTAFTVIGI 1152

Db    1109 qeassvceqsvnslpgsmakavayl-ekrlp-h----lt-npyavam--tsya-l-anag 1157
            : |  :|    |: :  :: ||  :| |: || :        |: ::||:::    : : : :  :
Qy    1153 RKAFDICPL-VK-IDTALIKADNFLLENTLPAQSTFTLAISAYALSLGDKTHPQFRSIVS 1210

Db    1158 klnketllkfasp------q-ldh--wpvpg-gyqytleatsyallalvkvkafeeagpi 1207
             |::|:|:| :|       : |:|    :||  |     :|:|:||||: :::| ::   |:
Qy    1211 ALKREALVKGNPPIYRFWKDNLQHKDSSVPNTGTARMVETTAYALLTSLNLKDINYVNPV 1270

Db    1208 vrwlnkqkkvgggygstqstimvfqavaeywshvkdlkdfdlninlevagrasvtkwsin 1267
            ::||: : : |||: ||| ||   :::::||    ||:|: : ::|::     :::: :  :
Qy    1271 IKWLSEEQRYGGGFYSTQDTINAIEGLTEYSLLVKQLR-LSMDIDVSYKHKGALHNYKMT 1329

Db    1268 nknqfhtrtdkvnsidkdltvkas-gngeatlsvvtlyyalpeekdsdcesfdlsvtltk 1326
            :|| |  |   |  :: || | :: |:| ||: | |:     :    :   | || | :
Qy    1330 DKN-FLGRPVEVL-LNDDLIVSTGFGSGLATVHVTTVVHKTSTSEEV-C-SFYLKIDTQD 1385

Db    1327 mdktshed-akesfmltie-vlykns--erda-tm-sildiglltgfivdtddlnqlskg 1380
            :: :     ::  ::  :     || |   | : :  :::||:| ||:  : :||: |  |
Qy    1386 IEASHYRGYGNSDYKRIVACASYKPSREESSSGSSHAVMDISLPTGISANEEDLKALVEG 1445

Db    1381 reryiekfemdkvlsergslilyldkvshkledrisfkihrvqevgvlqpaavsvyeyyn 1440
             :: :   ::: |    : | :|| |: ::          : |:|   : ||| | ||: :||||
Qy    1446 VDQLFTDYQI-K---D-GHVILQLNSIPSSDFLCVRFRIFELFEVGFLSPATFTVYEYHR 1500

Db    1441 -qkrcvkfyhpqreggtlsrlclgdvctcaeescs-mqkkgepdv--q-ridkacgagld 1495
             :|:| :|| : :    : ::|  | | |   |: ||    :  :  :  | : || : :
Qy    1501 PDKQCTMFY-STSNI-KIQKVCEGAACKCVEADCGQMQEELDLTISAETRKQTACKPEIA 1558

Db    1496 yvykatvvdsklthtdtytvkidlvi-kpgtdegvegknrdfmglaycrealglmqgkt 1554
            | || ::       :            |    : | | | |    :       : |:     | :|  |: |:
Qy    1559 YAYKVSITSITVENVFVKYKATL-LDIYKTGEAVAEKDSEITFIKKVTCTNA-ELVKGRQ 1616

Db    1555 ymimgksedlhrvedkgllqykyvlgeqtwieywpsqqectsrdyrevclgidefinqit 1614
            |:|||| | |: :   :  ::| | |      ||||||| :   |:|     :     :||| ::|
Qy    1617 YLIMGK-EALQ-IKYNFSFRYIYPLDSLTWIEYWPRDTTCSS--CQAFLANLDEFAEDIF 1672

Db    1615 tfgc 1618
              ||
Qy    1673 LNGC 1676
```

```
RESULT    13
ENTRY           C4HU       #type complete
TITLE           complement C4A precursor - human
CONTAINS        classical-complement-pathway C3/C5 convertase (EC 3.4.21.43)
                  C4b subunit; complement C4a anaphylatoxin
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            25-Feb-1985 #sequence_revision 23-Aug-1996 #text_change
                  16-Feb-1997
ACCESSIONS      I56095; A29177; B29177; A90845; A19311; A92337; S12866;
                  A17265; A32335; A27600; I58991; I37399; A01262; A01263
REFERENCE       I56095
   #authors     Yu, C.Y.
   #journal     J. Immunol. (1991) 146:1057-1066
   #title       The complete exon-intron structure of a human complement
                  component C4A gene. DNA sequences, polymorphism, and
                  linkage to the 21-hydroxylase gene.
```



```
   #cross-references MUID:91108039
   #accession    I56095
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues       1-1744 ##label RES
      ##cross-references GB:M59815; NID:g179672; CDS_PID:g179674
REFERENCE       A29177
   #authors     Belt, K.T.; Yu, C.Y.; Carroll, M.C.; Porter, R.R.
   #journal     Immunogenetics (1985) 21:173-180
   #title       Polymorphism of human complement component C4.
   #cross-references MUID:85156269
   #accession    A29177
      ##molecule_type DNA
      ##residues       1-22 ##label BEL
      ##cross-references GB:M14823
   #accession    B29177
      ##molecule_type DNA
      ##residues       1056-1225 ##label BE3
      ##cross-references GB:M14824
REFERENCE       A90845
   #authors     Belt, K.T.; Carroll, M.C.; Porter, R.R.
   #journal     Cell (1984) 36:907-914
   #title       The structural basis of the multiple forms of human
                  complement component C4.
   #cross-references MUID:84156544
   #accession    A90845
      ##molecule_type mRNA
      ##residues       20-346,'S',348-417,'A',419-725,'P',727-1200,'S',
                         1202-1285,'S',1287-1418,1422-1744 ##label BE2
      ##cross-references GB:K02403
REFERENCE       A19311
   #authors     Carroll, M.C.; Porter, R.R.
   #journal     Proc. Natl. Acad. Sci. U.S.A. (1983) 80:264-267
   #title       Cloning of a human complement component C4 gene.
   #cross-references MUID:83117835
   #accession    A19311
      ##molecule_type mRNA
      ##residues       1195-1285,'S',1287-1294 ##label CAR
      ##cross-references GB:J00080
REFERENCE       A92337
   #authors     Moon, K.E.; Gorski, J.P.; Hugli, T.E.
   #journal     J. Biol. Chem. (1981) 256:8685-8692
   #title       Complete primary structure of human C4a anaphylatoxin.
   #cross-references MUID:81264286
   #accession    A92337
      ##molecule_type protein
      ##residues       680-725,'PN',728-756 ##label MOO
REFERENCE       S12866
   #authors     Hessing, M.; van't Veer, C.; Hackeng, T.M.; Bouma, B.N.;
                  Iwanaga, S.
   #journal     FEBS Lett. (1990) 271:131-136
   #title       Importance of the alpha(3)-fragment of complement C4 for the
                  binding with C4b-binding protein.
   #cross-references MUID:91032049
   #accession    S12866
      ##molecule_type protein
      ##residues       757,'X',759-771;980-990 ##label HES
REFERENCE       A17265
   #authors     Campbell, R.D.; Gagnon, J.; Porter, R.R.
   #journal     Biochem. J. (1981) 199:359-370
   #title       Amino acid sequence around the thiol and reactive acyl groups
                  of human complement component C4.
   #cross-references MUID:82182029
   #accession    A17265
```

```
      ##molecule_type protein
      ##residues     957-1012,'E',1014-1044 ##label CAM
REFERENCE       A32335
   #authors     Chakravarti, D.N.; Campbell, R.D.; Porter, R.R.
   #journal     Mol. Immunol. (1987) 24:1187-1197
   #title       The chemical structure of the C4d fragment of the human
                  complement component C4.
   #cross-references MUID:88094444
   #accession    A32335
      ##molecule_type protein
      ##residues     957-1012,'E',1014-1108,'I',1110-1175,'S',1177-1270,'V',
                       1272-1336 ##label CHA
      ##note         1073-Gly, 1120-Leu, 1121-Ser, 1124-Ile, 1125-His,
                       1207-Ala, 1210-Arg were also found
REFERENCE       A27600
   #authors     Chakravarti, D.N.; Campbell, R.D.; Gagnon, J.
   #journal     FEBS Lett. (1983) 154:387-390
   #title       Amino acid sequence of a polymorphic segment from fragment
                  C4d of human complement component C4.
   #cross-references MUID:83158189
   #accession    A27600
      ##molecule_type protein
      ##residues     1199-1270,'V',1272-1299,'V',1301-1304 ##label CH2
REFERENCE       I58991
   #authors     Whitehead, A.S.; Goldberger, G.; Woods, D.E.; Markham, A.F.;
                  Colten, H.R.
   #journal     Proc. Natl. Acad. Sci. U.S.A. (1983) 80:5387-5391
   #title       Use of a cDNA clone for the fourth component of human
                  complement (C4) for analysis of a genetic deficiency of C4
                  in guinea pig.
   #cross-references MUID:83299979
   #accession    I58991
      ##status       translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues     1448-1474 ##label RE2
      ##cross-references GB:K00830; NID:g187772; CDS_PID:g187773
REFERENCE       I37396
   #authors     Sargent, C.A.; Anderson, M.J.; Hsieh, S.L.; Kendall, E.;
                  Gomez-Escobar, N.; Campbell, R.D.
   #journal     Hum. Mol. Genet. (1994) 3:481-488
   #title       Characterisation of the novel gene G11 lying adjacent to the
                  complement C4A gene in the human major histocompatibility
                  complex.
   #cross-references MUID:94282044
   #accession    I37399
      ##status       translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues     1-21 ##label RE3
      ##cross-references EMBL:X77491; NID:g453410; CDS_PID:g453411
COMMENT    This protein is synthesized as a single-chain precursor and, prior
             to secretion, is enzymatically cleaved to form a trimer of
             nonidentical chains (alpha, beta, and gamma) which are linked by
             disulfide bonds.
COMMENT    The activation of complement C4 by complement subcomponent C1s
             releases the C4a anaphylatoxin from the amino end of the alpha
             chain and generates C4b, which associates with the 2a fragment of
             complement factor 2 to form the classical-complement-pathway C3
             convertase. The C4b,C2a fragment then associates with the 3b
             fragment of complement factor 3 to form the
             classical-complement-pathway C5 convertase.
COMMENT    C4a anaphylatoxin is a vasoactive peptide and a mediator of
             inflammation.
COMMENT    The activity of C4b is regulated by proteolytic cleavage involving
             C4b-binding protein and factor I.
```



```
COMMENT    Residues 1446 or 1449 may be the carboxyl end of the alpha chain.
COMMENT    There are at least two genes coding for C4, C4A and C4B. Each gene
             has many alleles.
GENETICS
   #gene         GDB:C4A
      ##cross-references GDB:119732
   #map_position 6p21.3-6p21.3
   #introns      22/2; 88/3; 156/1; 179/3; 209/2; 237/1; 269/2; 304/3; 349/1;
                   387/3; 447/3; 508/3; 570/3; 623/3; 666/1; 691/1; 757/1;
                   794/2; 818/1; 864/3; 934/3; 952/1; 982/1; 1052/1; 1077/2;
                   1129/3; 1168/3; 1226/1; 1303/3; 1359/3; 1379/3; 1411/1;
                   1473/2; 1503/3; 1528/3; 1563/1; 1593/1; 1626/1; 1654/1;
                   1698/2
CLASSIFICATION   #superfamily alpha-2-macroglobulin
KEYWORDS         acute phase; complement classical pathway; glycoprotein;
                   hydrolase; inflammation; plasma; polymorphism; serine
                   proteinase; thiolester bond
FEATURE
   1-19              #domain signal sequence #status predicted #label SIG\
   20-675            #product complement C4 beta chain #status predicted
                       #label BET\
   20-675,757-1446,
   1454-1744         #product C4b #status predicted #label C4B\
   680-1446          #product complement C4 alpha chain #status predicted
                       #label ALP\
   680-756           #product C4a anaphylatoxin #status experimental #label
                       C4A\
   757-845           #region C4b-binding protein binding\
   957-1336          #product C4d fragment #status experimental #label C4D\
   1454-1744         #product complement C4 gamma chain #status predicted
                       #label GAM\
   756-757           #cleavage_site Arg-Ala (complement subcomponent C1s)
                       #status experimental\
   1010-1013         #cross-link thiolester (Cys-Gln) #status experimental\
   1328              #binding_site carbohydrate (Asn) (covalent) #status
                       experimental
SUMMARY          #length 1744  #molecular-weight 192860  #checksum 9431

  Query Match            14.0%;  Score 1692;  DB 2;  Length 1744;
  Best Local Similarity  28.2%;  Pred. No. 5.65e-280;
  Matches  466;  Conservative  419;  Mismatches 620;  Indels 147;  Gaps 116;

Db     139 ghlflqtdqpiynpgqrvryrvfaldqkmrpst-dtitvmvenshglrvrkkevympssi 197
           | ||::|| |:| | | |: ||::|:: ::|: :|:   :: : |  | : |     :|
Qy     124 GFLFIHTDKPVYTPDQSVKVRVYSLNDDLKPAKRETVLTFID-PEGSEVDMVEEIDHIGI 182

Db     198 fq-ddfvipdisepgtwkisarfsdglesnsstqfevkkyvlpnfevkitpgkpyiltvp 256
           :   || ||  :  | | | |:: : : : ::: |||| ||||:| | | |    :|
Qy     183 ISFPDFKIPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEPEYNFI-GYK 241

Db     257 ghldemqldiqaryiygkpv-qgvayvrfgllde--dgkktffrglesqtklvngqshis 313
           :: : :: | |||:| | | ::  |: ||: ::  |  |    :     |:|:|| ::::
Qy     242 -NFKNFEITIKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLINGIAQVT 300

Db     314 lskaefqdaleklnm-gitdlqglrlyvaaaiieypggemeeaeltswyfvsspfsldls 372
           :  :|   |:  |:  :: ||    ||:| ::|| || ||||: :  :| ||: |:|
Qy     301 FD-SE--TAVKELSYYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLNLV 357

Db     373 ktkrhlvpgapfllqalvremsgspasgipvkvsa-tvsspgsvpevqdiqqntd-gsgq 430
            |   | || |: :   |::       :|:|| ::| |:       ::::    |    |
Qy     358 ATPLFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDDGV 417

Db     431 vsipiiipqtiselqlsvsagsph-pai--arltvaapp-sg-gpgflsierpds-rppr 484
           |: : :|  :: |:::| : :|  |     ||    | : |: : ::| |:: |: ::
```

```
Qy    418 ASFVLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKALL 477

Db    485 vgdtlnlnlravgs-gatfshyyymilsrgqiv-fmnre--pkrtltsvsvfvdhhlaps 540
          ||: ||: : : ::    ::|| |:|||:| |: |  ||  :  :  |::: | ::: ||
Qy    478 VGEHLNIIVTPKSPYIDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQNMVPS 537

Db    541 fyfvafyy-hgdh-p--vanslrvdvqagacegklelsvd-gakqyrngesvklhletds 595
            :: :|   |:: :  |::|::::::   |    |:: :    |  |   |::| |:: |
Qy    538 SRLLVYYIVTGEQTAELVSDSVWLNIEE-KCGNQLQVHLSPDADAYSPGQTVSLNMATGM 596

Db    596 lalvalgaldtalyaagskshkplnmgkvfeamnsydlgcgpgggdsalqvfqaaglaf- 654
           : |||:|:|:|:|:    : |||:  :||: ::  |||||:|||  :    ||: |||:|
Qy    597 DSWVALAAVDSAVYGVQRGAKKPLE--RVFQFLEKSDLGCGAGGGLNNANVFHLAGLTFL 654

Db    655 sdgdqwtlsrkrlscpkekttrkkrnvnfqkaineklgqyasptakrccqdgvtrlpmmr 714
          ::::     . :   :| ||    | :| :   || |:|   : |  :   |:|| || : :
Qy    655 TNANADDSQENDEPC-KE-ILRPRRTL--QKKIEEIAAKYKHSVVKKCCYDG-ACVNNDE 709

Db    715 sceqraarvqqld-crepflsccqfaeslrkksrdkgqaglqraleilqeedlideddip 773
          :||||||:      |  :|  ||  |  ||  :  :  ::   :|   |   |:   |:
Qy    710 TCEQRAARISLGPRCIKAFTECCVVASQLR--A-NISHKDMQ--LGRLHMKTLLPVSKPE 764

Db    774 vrsffpenwlwrvetvdrfqiltlwlpdslttweihglslsktkglcvatpvqlrvfref 833
          :||:|||:|||  |  | |   |  : ||||||||||:|:::|:| |:||| |  :||::
Qy    765 IRSYFPESWLWEVHLVPRRKQLQFALPDSLTTWEIQGIGISNT-GICVADTVKAKVFKDV 823

Db    834 hlhlrlpmsvrrfeqlelrpvlynyldknltvsvhvspveglclagggglaqqvlvpags 893
            | : :| || | ||::|:  :|||   |    :  | :|:|||:| : :  : :|   ::
Qy    824 FLEMNIPYSVVRGEQIQLKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTKSSKC 883

Db    894 arp-vafsvvptaaaav-slkvvarg-sfefpv--g-davskvlqiekegaihreelvye 947
           |  |  |         :  :| :|: :  :  |:  :   |  : :  || : ||   |
Qy    884 VRQKVEGSSSHLVTFTVLPLEIGLHNINFSLETWFGKEILVKTLRVVPEG-VKRES--YS 940

Db    948 lnpldhrgrtleipgnsd-pnmip-dgdfnsyv-rvtasdpldtlgse-ga-lspggvas 1002
              || ||    |:   : :|  :|  | ||    :: : |: :    |  :|   :| || |:
Qy    941 GVTLDPRGIYGTISRRKEFPYRIPLDLVPKTEIKRILSVKGL-LVGEILSAVLSQEGINI 999

Db   1003 llrlprgcgeqtmiylaptlaasryldkteqwstlppet-kd-havdliqk-gymriqqf 1059
          | :||:| :|   :: : | :   :||:  ::|: : ::    : : :   | | : | :
Qy   1000 LTHLPKGSAEAELMSVVPVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEGMLSIMSY 1059

Db   1060 rkadgsyaawlsrdsstwltafvlkvlslaqeqvggspeklqetsnwllsq-qqadgsfq 1118
          |:|| ||: |·:  :|||||||  |:||:     |  : : : ::  ||:   |  :|||
Qy   1060 RNADYSYSVWKGGSASTWLTAFALRVLGQVNKYVEQNQNSICNSLLWLVENYQLDNGSFK 1119

Db   1119 dpcpvldrsmqggl-v-gndetvaltafvtialhhglavfqdegaeplkqrveasiskan 1176
          :         :|| |  | : :::: ||||  |:::    | | |    ||  :::::: ||:
Qy   1120 ENSQYQPIKLQGTLPVEARENSLYLTAFTVIGIRK--A-F-DI-C-PLV-KIDTALIKAD 1172

Db   1177 sflgekasagllgahaaaitayaltlt-kapvdllgvahnnlmamaqetgdn-lywgsvt 1234
          :|| |:: ::  :: : ||:||||:|  |:  :: ::  :  : |:  |    |:  :|:
Qy   1173 NFLLENTLPAQ-STFTLAISAYALSLGDKTHPQFRSIV-SALKREALVKGNPPIYRFWKD 1230

Db   1235 gsqsnavsptpaprnpsdpmpqapalwiettayallhlllhegkaemadqaaawltrqgs 1294
            | :  |  :  | | :      |  : :||||||||   | :       :   :    ||: :
Qy   1231 NLQHKD-S-S-VP-N-TGT---AR-M-VETTAYALLTSL-NLKDINYVNPVIKWLSEEQR 1279

Db   1295 fqggfrstqdtvialdalsaywiashtteerglnvtlsstgrngfkshalqlnnrqirgl 1354
           : ||| |||||: |:::|: | :      : | |:: :    :    :| |   |:|   : :
Qy   1280 YGGGFYSTQDTINAIEGLTEYSLL--VKQLR-LSMDID-VS---YK-HKGALHN--YK-M 1328

Db   1355 eeelqfslgskinvkvggnskgtlkvlrtynvldmkknttcqdlqievtvkghveytmean 1414
            :  | ||  ::| :  |     | |   :    :  :|   :::  ||  |   |  ::
```



```
Qy   1329 TDK-NF-LGRPVEVLL--ND-D-LIVSTGFGS-GL--AT---VHV-TTV-VH-K-T-STS 1371

Db   1415 edyedyeydelpakddpdaplqpvtplqlfegrrnrrrreapkvveeqesrvhytvciwr 1474
          |:   : |  |  | | :  ::: :  : : |   : :|      :|        |     :
Qy   1372 EEVCSF-Y--L--KIDTQD-IEA-SHYRGY-GNSDYKR-----IVA---C-ASYKPS-RE 1413

Db   1475 ngkvglsgmaiadvtllsgfhalradlekltslsdryvshfetegphvllyfdsvptsr- 1533
          :: | |: |: |::| :|: |   || |    |:  : ::     ||:| ::|:|:|
Qy   1414 ESSSG-SSHAVMDISLPTGISANEEDLKALVEGVDQLFTDYQIKDGHVILQLNSIPSSDF 1472

Db   1534 ecvgfeavqevpvglvqpasatlydyynperrcsvfygapsksrllatlcsaevcqcaeg 1593
            || |   :   ||:: ||: |:|:|  |:::|::||:: |: ::    :| :  | | |:
Qy   1473 LCVRFRIFELFEVGFLSPATFTVYEYHRPDKQCTMFYST-SNIKI-QKVCEGAACKCVEA 1530

Db   1594 kcprqrralerglqdedgyrmkfacyyprveygfqvkviredsraafrlfetkitqvlhf 1653
            | : :  |:  :   :  |:  ||  | : |:: | :          |  : : : ::
Qy   1531 DCGQMQEELDLTI-SAET-RKQTACK-PEIAYAYKVSITSITVENVFVKYKATLLDIYKT 1587

Db   1654 tkdvkaaanqmrnflvrasc-rlrlepgkeylimgldg--atydleghpqylldsnswie 1710
              |     :::  |: : :|    |  |::||||| ::     |:: :   | ||| :|||
Qy   1588 GEAVAEKDSEI-TFIKKVTCTNAELVKGRQYLIMGKEALQIKYNFSFRYIYPLDSLTWIE 1646

Db   1711 empserlcrstrqraacaqlndflqeygtqgc 1742
              | :  | |: | |  | |::| ::    ||
Qy   1647 YWPRDTTC-SSCQ-AFLANLDEFAEDIFLNGC 1676
```

RESULT 14
ENTRY A24558 #type complete
TITLE complement C4 precursor - mouse
CONTAINS classical-complement-pathway C3/C5 convertase (EC 3.4.21.43) C4b subunit; complement C4a anaphylatoxin
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 31-Mar-1989 #sequence_revision 11-Nov-1994 #text_change 13-Mar-1997
ACCESSIONS A24558; A25371; A21692; A30520; A60227; A22039; A29059; A01264; B41195; I59084; I48274; I54567; I69023
REFERENCE A24558
#authors Sepich, D.S.; Noonan, D.J.; Ogata, R.T.
#journal Proc. Natl. Acad. Sci. U.S.A. (1985) 82:5895-5899
#title Complete cDNA sequence of the fourth component of murine complement.
#cross-references MUID:85298264
#accession A24558
##molecule_type mRNA
##residues 1-1738 ##label SEP
##cross-references GB:M11729
##experimental_source strain B10.WR
REFERENCE A25371
#authors Nonaka, M.; Nakayama, K.; Yeul, Y.D.; Takahashi, M.
#journal J. Biol. Chem. (1985) 260:10936-10943
#title Complete nucleotide and derived amino acid sequences of the fourth component of mouse complement (C4). Evolutionary aspects.
#cross-references MUID:85289294
#accession A25371
##molecule_type mRNA
##residues 1-131,'Y',133-326,'E',328-569,'E',571-1323,'N', 1325-1441,'K',1443-1452,'V',1454-1738 ##label NON
##cross-references GB:M11789
##experimental_source strain FM
REFERENCE A94013
#authors Nonaka, M.; Takahashi, M.; Natsuume-Sakai, S.; Nonaka, M.; Tanaka, S.; Shimizu, A.; Honjo, T.

```
   #journal      Proc. Natl. Acad. Sci. U.S.A. (1984) 81:6822-6826
   #title        Isolation of cDNA clones specifying the fourth component of
                   mouse complement and its isotype, sex-limited protein.
   #cross-references MUID:85038607
   #accession    A21692
     ##molecule_type mRNA
     ##residues      651-719,'G',721-738,'AI',741-805 ##label NO2
     ##cross-references GB:M12970
     ##experimental_source strain FM
REFERENCE        A30520
   #authors      Taillon-Miller, P.A.; Shreffler, D.C.
   #journal      J. Immunol. (1988) 141:2382-2387
   #title        Structural basis for the C4d.1/C4d.2 serologic allotypes of
                   murine complement component C4.
   #cross-references MUID:89009745
   #accession    A30520
     ##molecule_type DNA
     ##residues      961-1205,'Q',1207-1290 ##label TAI
     ##cross-references GB:M23186
     ##experimental_source strain B10.BR
REFERENCE        A60227
   #authors      Ogata, R.T.; Zepf, N.E.
   #journal      Eur. J. Immunol. (1990) 20:1607-1610
   #title        C4 from C4-high and C4-low mouse strains have identical
                   sequences in the region corresponding to the
                   isotype-specific segment of human C4.
   #cross-references MUID:90353398
   #accession    A60227
     ##molecule_type DNA
     ##residues      1099-1142 ##label OGA
     ##cross-references GB:X55493
REFERENCE        A22039
   #authors      Levi-Strauss, M.; Tosi, M.; Steinmetz, M.; Klein, J.; Meo, T.
   #journal      Proc. Natl. Acad. Sci. U.S.A. (1985) 82:1746-1750
   #title        Multiple duplications of complement C4 gene correlate with
                   H-2-controlled testosterone-independent expression of its
                   sex-limited isoform, C4-Slp.
   #cross-references MUID:85166208
   #accession    A22039
     ##molecule_type mRNA
     ##residues      1105-1118,'A',1120-1189,'T',1191-1449 ##label LEV
     ##cross-references GB:K02798
     ##experimental_source strain B10.W7R
REFERENCE        A93753
   #authors      Tosi, M.; Levi-Strauss, M.; Duponchel, C.; Meo, T.
   #journal      Philos. Trans. R. Soc. Lond. (1984) 306:389-394
   #title        Sequence heterogeneity of murine complementary DNA clones
                   related to the C4 and C4-Slp isoforms of the fourth
                   complement component.
   #accession    A29059
     ##molecule_type mRNA
     ##residues      1258-1376 ##label TOS
     ##cross-references GB:K02798
REFERENCE        A01264
   #authors      Ogata, R.T.; Shreffler, D.C.; Sepich, D.S.; Lilly, S.P.
   #journal      Proc. Natl. Acad. Sci. U.S.A. (1983) 80:5061-5065
   #title        cDNA clone spanning the alpha-gamma subunit junction in the
                   precursor of the murine fourth complement component (C4).
   #cross-references MUID:83273751
   #accession    A01264
     ##molecule_type mRNA
     ##residues      1360-1400,'S',1402-1511 ##label OG2
     ##cross-references GB:K00019
     ##experimental_source strain B10.W7R
```



```
REFERENCE        A41195
   #authors      Ogata, R.T.; Sepich, D.S.
   #journal      Proc. Natl. Acad. Sci. U.S.A. (1984) 81:4908-4911
   #title        Genes for murine fourth complement component (C4) and
                   sex-limited protein (Slp) identified by hybridization to
                   C4- and Slp-specific cDNA.
   #cross-references MUID:84272739
   #accession    B41195
     ##molecule_type mRNA
     ##residues      1360-1400,'S',1402-1511 ##label OG3
     ##cross-references GB:K00019
     ##experimental_source strain B10.W7R
REFERENCE        I59084
   #authors      Nonaka, M.; Kimura, H.; Yeul, Y.D.; Yokoyama, S.; Nakayama,
                   K.; Takahashi, M.
   #journal      Proc. Natl. Acad. Sci. U.S.A. (1986) 83:7883-7887
   #title        Identification of the 5'-flanking regulatory region
                   responsible for the difference in transcriptional control
                   between mouse complement C4 and Slp genes.
   #cross-references MUID:87017050
   #accession    I59084
     ##status        preliminary; translated from GB/EMBL/DDBJ
     ##molecule_type DNA
     ##residues      1-21 ##label RES
     ##cross-references GB:M14225; NID:g199291; CDS_PID:g554211
REFERENCE        I48274
   #authors      Hemenway, C.; Kalff, M.; Stavenhagen, J.; Walthall, D.;
                   Robins, D.
   #journal      Nucleic Acids Res. (1986) 14:2539-2554
   #title        Sequence comparison of alleles of the fourth component of
                   complement (C4) and sex-limited protein (Slp).
   #cross-references MUID:86176748
   #accession    I48274
     ##status        preliminary; translated from GB/EMBL/DDBJ
     ##molecule_type mRNA
     ##residues      591-603,'M',605-1323,'N',1325-1452,'V',1454-1585,'Q',
                       1587-1738 ##label RE2
     ##cross-references EMBL:X05314; NID:g50241; CDS_PID:g50242
REFERENCE        I54567
   #authors      Nonaka, M.; Nakayama, K.; Yeul, Y.D.; Shimizu, A.; Takahashi,
                   M.
   #journal      Immunol. Rev. (1985) 87:81-99
   #title        Molecular cloning and characterization of complementary and
                   genomic DNA clones for mouse C4 and SLP.
   #cross-references MUID:86031969
   #accession    I54567
     ##status        preliminary; translated from GB/EMBL/DDBJ
     ##molecule_type DNA
     ##residues      1-128 ##label RE3
     ##cross-references GB:M12968; NID:g199267; CDS_PID:g199270
   #accession    I69023
     ##status        preliminary; translated from GB/EMBL/DDBJ
     ##molecule_type DNA
     ##residues      1724-1738 ##label RE4
     ##cross-references GB:M12969; NID:g199268; CDS_PID:g387439
COMMENT      This protein is synthesized as a single-chain precursor and, prior
               to secretion, is enzymatically cleaved to form a trimer of
               nonidentical chains (alpha, beta, and gamma) which are linked by
               disulfide bonds.
COMMENT      The activation of complement C4 by complement subcomponent C1s
               releases the C4a anaphylatoxin from the amino end of the alpha
               chain and generates C4b, which associates with the 2a fragment of
               complement factor 2 to form the classical-complement-pathway C3
               convertase. The C4b,C2a fragment then associates with the 3b
```

```
            fragment of complement factor 3 to form the
            classical-complement-pathway C5 convertase.
COMMENT     C4a anaphylatoxin is a vasoactive peptide and a mediator of
            inflammation.
COMMENT     The activity of C4b is regulated by proteolytic cleavage involving
            C4b-binding protein and factor I.
GENETICS
   #introns        22/2; 86/3; 977/3; 1047/3; 1073/3; 1125/3; 1164/3; 1221/3
   #note           the list of introns is incomplete
CLASSIFICATION     #superfamily alpha-2-macroglobulin
KEYWORDS           acute phase; complement classical pathway; glycoprotein;
                     hydrolase; inflammation; plasma; polymorphism; serine
                     proteinase; thiolester bond
FEATURE
   1-19               #domain signal sequence #status predicted #label SIG\
   20-673             #product complement C4 beta chain #status predicted
                        #label BET\
   20-673,754-1440,
   1448-1738          #product complement C4b #status predicted #label C4B\
   678-1440           #product complement C4 alpha chain #status predicted
                        #label ALP\
   678-753            #product C4a anaphylatoxin #status predicted #label C4A\
   754-843            #region C4b-binding protein binding\
   953-1332           #product C4d fragment #status predicted #label C4D\
   1448-1738          #product complement C4 gamma chain #status predicted
                        #label GAM\
   224,743,1387       #binding_site carbohydrate (Asn) (covalent) #status
                        predicted\
   753-754            #cleavage_site Arg-Asn (complement subcomponent C1s)
                        #status predicted\
   1006-1009          #cross-link thiolester (Cys-Gln) #status predicted
SUMMARY            #length 1738  #molecular-weight 192870  #checksum 4149

  Query Match            13.3%;  Score 1605;  DB 2;  Length 1738;
  Best Local Similarity  27.6%;  Pred. No. 1.46e-263;
  Matches  449;  Conservative  407;  Mismatches  628;  Indels 143;  Gaps 111;

Db     137 ghifvqtdqpiynpgqrvryrvfaldqkmrpstdfltitvenshglrv-lkkei-ftsts 194
           | :|::|| |:| | | |: ||::|:: ::|:      :|   :: |   | :  ||    :
Qy     124 GFLFIHTDKPVYTPDQSVKVRVYSLNDDLKPAKRETVLTFIDPEGSEVDMVEEIDHIGII 183

Db     195 ifqdaftipdisepgtwkisarfsdglesnrsthfevkkyvlpnfevkitpwkpyilmvp 254
           | | | ||  :   | | | |:: : : :  :: |||| ||||:| | | |   | ::
Qy     184 SFPD-FKIPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEP-E-YNFIGY 240

Db     255 snsdeiqldiqaryiygkpv-qgvaytrfalmde--qgkrtflrgletqaklvegrthis 311
           |   :::: | |||:| | | ::  |  |:: :: :  :  :  ::     ::|::| ::::
Qy     241 KNFKNFEITIKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLINGIAQVT 300

Db     312 iskdqfqaaldkinigvrdleglrlyaataviespggemeeaeltswrfvssafsldlsr 371
            :  :       |        :: ||:   || | :|||| ||   ||||: : ::| |:: |:|
Qy     301 FDSETAVKELS--YYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLNLVA 358

Db     372 tkrhlvpgahfllqalvqemsgseasnvpvkvsa-tlvsgsdsqvldiqqstngiqq-v- 428
            |   | ||  : :    | :   :   : ||| ::| |:     ::  ||    |    : : |
Qy     359 TPLFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDDGVA 418

Db     429 sisfpipptvtelrllvsagsl-ypai--arltvqa-p-psrgtgflsie-pldprspsv 482
           |: : :|: || | : | : :      |        ||    :| : :| : ::| |:    : :: |
Qy     419 SFVLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKALLV 478

Db     483 qdtfilnlqpvgipaptfshyyymiisrgqimamq-rep-rktv-tsvsvlvdhqlapsf 539
           |: : : : | :     ::|| |:|:|:| |:  | ||     :    |::: | : : ||
Qy     479 GEHLNIIVTPKSPYIDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQNMVPSS 538
```



```
Db     540 yfvayfyhqg-h-p--vansllliniqsrdcegklqlkvd-gakeyrnadmmklriqtdsk 594
            :: |:    | : :  |::|: :||: : |    ||: :     |  |  :: : | : |
Qy     539 RLLVYYIVTGEQTAELVSDSVWLNIEEK-CGNQLQVHLSPDADAYSPGQTVSLNMATGMD 597

Db     595 alvalgavdtalyavggrshkpldmskvfevinsynvgcgpgggddalqvfqdaglafsd 654
            : |||:|||:|:|:|    : |||:   :||: ::  ::|||:||| :    ||: |||:|
Qy     598 SWVALAAVDSAVYGVQRGAKKPLE--RVFQFLEKSDLGCGAGGGLNNANVFHLAGLTFLT 655

Db     655 gdrltqtre-dlscpkekksrqkrnvnfqkavseklgqysspdakrccqdgmtklpmkrt 713
                :::| | :| ||   | :| :   || : |  : |   :  |:|| ||    :     |
Qy     656 NANADDSQENDEPC-KEIL-RPRRTL--QKKIEEIAAKYKHSVVKKCCYDGAC-VNNDET 710

Db     714 ceqraarvpqqa-crepflscckfaedlrrnqtrsqahlarnnhnmlqeedlideddilv 772
           ||||||:: :   : |  :|   ||   | :|| | :: : :|:|   | |        |:    :
Qy     711 CEQRAARISLGPRCIKAFTECCVVASQLRANISHKDMQLGRL-H-M-K--TLLPVSKPEI 765

Db     773 rtsfpenwlwrvepvdssklltvwlpdsmttweihgvslskskglcvakptrvrvfrkfh 832
           |: |||:||| |    |    | |    ||||:||||:|:::|:: |:|||     : :||:
Qy     766 RSYFPESWLWEVHLVPRRKQLQFALPDSLTTWEIQGIGISNT-GICVADTVKAKVFKDVF 824

Db     833 lhlrlpisirrfeqfelrpvlynylnddvavsvhvtpveglclagggmmaqqvtvpagsa 892
           | : :| |: | ||::|:  :||   :      | :::|||:| : : :: :| | ::
Qy     825 LEMNIPYSVVRGEQIQLKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTKSSKCV 884

Db     893 rp-vafsvvptaaanv-plkvvargv-fdl----g-davskilqiekegaihreelv-yn 943
           |   |   |       :  | ||  :   : : | |    | : : | |::  || : ||
Qy     885 RQKVEGSSSHLVTFTVLPLEIGLHNINFSLETWFGKEILVKTLRVVPEG-VKRESYSGVT 943

Db     944 ldp--l-nnlgrtleipgssdpnivpdgdfsslvrvtasepletmgsegalspggvasll 1000
           |||  :   ::| |:|       ::|| ::  :: | :     | : | : || |:  |
Qy     944 LDPRGIYGTISRRKEFPYRIPLDLVPKTEIKRILSVKGLLVGEIL-S-AVLSQEGINILT 1001

Db    1001 rlpqgcaeqtmiylaptltasnyldrteqwsklspet-kd-havdliqk-gymriqqfrk 1057
            :|| | ||   :: : | :    :||:  ::|: : ::    : : :       | | : |  :|:
Qy    1002 HLPKGSAEAELMSVVPVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEGMLSIMSYRN 1061

Db    1058 ndgsfgawlhrdsstwltafvlkilslaqeqvgnspeklqetaswllaq-qlgdgsfhdp 1116
             | |:: |    :||||||| |::|:       |  : : : :: ||:   || :||| :
Qy    1062 ADYSYSVWKGGSASTWLTAFALRVLGQVNKYVEQNQNSICNSLLWLVENYQLDNGSFKEN 1121

Db    1117 cpvihramqggl-vgs-detvaltafvvialhhgldvfqdddakqlknrveasitkansf 1174
                :|| | | : :::: |||| ||::: :     |    |       | ::::::  ||::|
Qy    1122 SQYQPIKLQGTLPVEARENSLYLTAFTVIGIRKA---F--DIC-PL-VKIDTALIKADNF 1174

Db    1175 lgqkasagllgahaaaitayaltlt-kasedlrnvahnslmamaeetgeh-lywglvlgs 1232
           | ::: ::   :: : ||:||||:|   |:   ::|::  ::|:  |    |:   :|:    :
Qy    1175 LLENTLPAQ-STFTLAISAYALSLGDKTHPQFRSIV-SALKREALVKGNPPIYR---F-W 1228

Db    1233 qdkvvlrptaprsptepvpqapalwiettayall-hlllregkgkmadkaaswlthqgsf 1291
           |:  |:   :   | :      |  : :||||||| | |::    :  :       ||: :  :
Qy    1229 KDN--LQHKDSSVP-NTGT-AR-M-VETTAYALLTSLNLKD-I-NYVNPVIKWLSEEQRY 1280

Db    1292 hgafrstqdtvvtldalsaywiashtteekalkvtlssmgrnglkthglhlnnhqvkgle 1351
           |:| |||||: ::::|: |    |        | |:::: :  :  | :| |:|  |:   |  :
Qy    1281 GGGFYSTQDTINAIEGLTEY---SLLV--KQLRLSMD-ID-VSYKHKGA-LHNY--K-MT 1329

Db    1352 eelkfslgstisvkvegnskgtlkilrtynvldmknttcqdlqievkvtgaveyawdane 1411
             : :| || : | :   |          | :    :     | | || | :  ::|
Qy    1330 DK-NF-LGRPVEVLL--ND-D-L-IVST-G-FG----S--GLAT-VHVTTVVHKT-STSE 1372

Db    1412 dyedyydmpaaddpsvplqpvtplqlfegrrsrrrreapkvaeeqesrvqytvciwrngk 1471
            :   :| :    |    :  :: :  : : |  | :|      ||          |     : ::
Qy    1373 EVCSFY-LKI-DTQDI--EA-SHYRGY-GN-SDYKR-I--VA----C-ASYKPS-REESS 1416
```

```
Db   1058 sfgawlhrdsstwltafvlkilslaqeqvgnspeklqetaswllaq-qlgdgsfhdpcpv 1116

Qy   1065 SYSVWKGGSASTWLTAFALRVLGQVNKYVEQNQNSICNSLLWLVENYQLDNGSFKENSQY 1124

Db   1117 ihramqggl-vgsne-tvaltafvvialhhglnvfreghakqlknrveasitkansflgq 1174

Qy   1125 QPIKLQGTLPVEARENSLYLTAFTVIGIR---KAF-DICP--L-VKIDTALIKADNFLLE 1177

Db   1175 kasagllgahaaaitayaltlt-kasedlrnvahnslmamaeetgen-lywglvlgsqdk 1232

Qy   1178 NTLPAQ-STFTLAISAYALSLGDKTHPQFRSIV-SALKREALVKGNPPIYR---F-WKDN 1231

Db   1233 vmlrpadprsptepvpqapalwiettayall-hlllregkgkmadkaaswlthqgnfhga 1291

Qy   1232 --LQHKDSSVP-NTGT-AR-M-VETTAYALLTSLNLKD-I-NYVNPVIKWLSEEQRYGGG 1283

Db   1292 frstqdtvvtldalsaywiashtteekalnvtlssmgrnglkthglhlnnhqvkgleeel 1351

Qy   1284 FYSTQDTINAIEGLTEY---SLLV--KQLRLSMD-ID-VSYKHKGA-LHNY--K-MTDK- 1331

Db   1352 kfslgstisvkvegnskgtlkilrtynvldmknttcqdlqievkvtdaveyawsayedye 1411

Qy   1332 NF-LGRPVEVLL--ND-D-L-IVST-G-FG----S--GLAT-VHVTTVVHKT-STSEE-V 1374

Db   1412 ddynmpatddpsvplqpvtplqlfegrrsrrreapkvaeeresrvhytvciwhngklgl 1471

Qy   1375 CSFYLK-IDTQDI--EA-SHYRGY-GN-SDYKR-I--VA--CAS---YKPS-REESSSG- 1418

Db   1472 sgmaiaditllsgfhalradlekltslsdryvshfetdgphvllyfdsvpttr-ecvgfg 1530

Qy   1419 SSHAVMDISLPTGISANEEDLKALVEGVDQLFTDYQIKDGHVILQLNSIPSSDFLCVRFR 1478

Db   1531 asqevvvglvqpasavlydyyspdhkcsvfyaaptksqllatlcsgdvcqcaegkcprll 1590

Qy   1479 IFELFEVGFLSPATFTVYEYHRPDKQCTMFYST-SNIKI-QKVCEGAACKCVEADCGQMQ 1536

Db   1591 rslerrvedkdgyrmrfacyyhqveygftvkvlredgraafrlfeskitqvlhfrt-dtm 1649

Qy   1537 EELDLTISA-ET-RKQTACK-PEIAYAYKVSITSITVENVFVKYKATLLDI--YKTGEAV 1591

Db   1650 asigqtrnflsrasc-rlrlepnkeylimgmdgetsdnkgdpqyl--ldsntwieempse 1706

Qy   1592 AEKDSEITFIKKVTCTNAELVKGRQYLIMGKEALQIKYNFSFRYIYPLDSLTWIEYWPRD 1651

Db   1707 qmcks 1711

Qy   1652 TTCSS 1656
```

Search completed: Wed Jan 28 12:15:48 1998
Job time : 99 secs.

```
Db   1472 lglsgmaiaditllsgfhalradlekltslsdryvshfetdgphvllyfdsvpttr-ecv 1530
          | |: |: ||:| :|: |   || |    |: : ::    ||:| ::|:|::   ||
Qy   1417 SG-SSHAVMDISLPTGISANEEDLKALVEGVDQLFTDYQIKDGHVILQLNSIPSSDFLCV 1475

Db   1531 gfgasqevvvglvqpssavlydyyspdhkcsvfyaaptksqllatlcsgdvcqcaegkcp 1590
          |   :   ||:: |:: :|:|  ||  |::||:: :: :   :| |  | | |: |
Qy   1476 RFRIFELFEVGFLSPATFTVYEYHRPDKQCTMFYST-SNIKI-QKVCEGAACKCVEADCG 1533

Db   1591 rllrslerrvedkdgyrmrfacyyprveygftvkvlredgraafrlfeskitqvlhfrkd 1650
          ::   |: :   : |:: || | : |:: | :         | : : : ::     :
Qy   1534 QMQEELDLTISA-ET-RKQTACK-PEIAYAYKVSITSITVENVFVKYKATLLDIYK-TGE 1589

Db   1651 tmasigqtrnflsrasc-rlrlepnkeylimgmdgetsdnkgdpqyl--ldsntwieemp 1707
          ::|       |: : :|   |   ::||||:::      :   :|:  ||| ||||  |
Qy   1590 AVAEKDSEITFIKKVTCTNAELVKGRQYLIMGKEALQIKYNFSFRYIYPLDSLTWIEYWP 1649

Db   1708 seqmcks 1714
            : | |
Qy   1650 RDTTCSS 1656
```

RESULT 15

```
ENTRY           S54784     #type complete
TITLE           sex-limited protein Slp(w7) - mouse
ORGANISM        #formal_name Mus musculus #common_name house mouse
DATE            27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change
                  13-Mar-1997
ACCESSIONS      S54784; I48770
REFERENCE       S54784
   #authors     Ogata, R.T.; Zepf, N.E.
   #journal     J. Immunol. (1991) 147:2756-2763
   #title       The murine Slp gene. Additional evidence that sex-limited
                  protein has no biologic function.
   #accession   S54784
      ##status        preliminary; translation not shown
      ##molecule_type DNA
      ##residues      1-1735 ##label OGA
      ##cross-references EMBL:M64933
REFERENCE       I48274
   #authors     Hemenway, C.; Kalff, M.; Stavenhagen, J.; Walthall, D.;
                  Robins, D.
   #journal     Nucleic Acids Res. (1986) 14:2539-2554
   #title       Sequence comparison of alleles of the fourth component of
                  complement (C4) and sex-limited protein (Slp).
   #cross-references MUID:86176748
   #accession   I48770
      ##status        preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues      634-641,'L',643-828 ##label RES
      ##cross-references EMBL:X06455; NID:g54102; CDS_PID:g899336
GENETICS
   #introns     22/2; 86/3; 154/1; 177/3; 207/2; 235/1; 267/2; 302/3; 347/1;
                  385/3; 444/3; 506/3; 568/3; 621/3; 664/1; 689/1; 754/1;
                  789/2; 813/1; 859/3; 927/3; 945/1; 975/1; 1045/1; 1070/2;
                  1122/3; 1161/3; 1219/1; 1296/3; 1352/3; 1372/3; 1404/1;
                  1464/2; 1494/3; 1519/3; 1554/1; 1584/1; 1617/1; 1645/1;
                  1689/2
CLASSIFICATION  #superfamily alpha-2-macroglobulin
SUMMARY         #length 1735  #molecular-weight 192820  #checksum 3834

  Query Match             13.0%;  Score 1565;  DB 6;  Length 1735;
  Best Local Similarity   27.2%;  Pred. No. 5.00e-256;
  Matches  442;  Conservative  420;  Mismatches  621;  Indels  142;  Gaps  111;
```



```
Db    137 ghifvqtdqpiynpgqpvryrvfaldqkmrpstdfltitvenshglrv-lkkei-ftsts 194
          | :|::|| |:| | |:|: ||::|:: ::|:     :|  :: |  | :  ||    :
Qy    124 GFLFIHTDKPVYTPDQSVKVRVYSLNDDLKPAKRETVLTFIDPEGSEVDMVEEIDHIGII 183

Db    195 ifqdaftipdisepgtwkisarfsdglesnrsthfevkkyvlpnfevkitpwkpyilmvp 254
          | | | ||  :  | | | |:: : : :  :: |||| ||||:| | | |     | ::
Qy    184 SFPD-FKIPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEP-E-YNFIGY 240

Db    255 snsdeiqldiqaryiygkpv-qgvaytrfalmde--qgkrtflrgletqaklvegrthis 311
          |  :::: | |||:| | | ::  |  |:: ::  :  :  ::     ::|::| ::::
Qy    241 KNFKNFEITIKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLINGIAQVT 300

Db    312 iskdqfqaaldkinigvrdleglrlyaataviespggemeeaeltswrfvssafsldlsr 371
          :  :      |      :: ||:  || | :|||| ||  ||||: : ::| |:: |:|
Qy    301 FDSETAVKELS--YYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLNLVA 358

Db    372 tkrhlvpgahfllqalvreisgseasnvivkvsa-tlvsgsdsqvlnvqqstnrigq-v- 428
          |    | ||  : :   |::        : | | ::| |:      ::  |:   |  |: : |
Qy    359 TPLFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDDGVA 418

Db    429 sisfpipptvtelrllvsagsl-yp--tiarltvqs-p-psrgtgflsie-pldprsprv 482
          |: : :|: || | : | : :    |      ||   :: : : :| : ::| |:   : ::  |
Qy    419 SFVLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKALLV 478

Db    483 gdtfilnlqavgipaptfshyyymiisrgqimam-srearrtvt--svsvlvdhqlapsf 539
          |: : : : : :     ::|| |:|:|:| |:   :||     :  |::: | : : ||
Qy    479 GEHLNIIVTPKSPYIDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQNMVPSS 538

Db    540 yfvayfyhqg-h-p--vansllliniqprdcegklqlkvv-gakeyhngdmmklqiqtdsk 594
          :: |:   | : :   |::|: :||: : |   ||: :    |   |   |: : | : |
Qy    539 RLLVYYIVTGEQTAELVSDSVWLNIEEK-CGNQLQVHLSPDADAYSPGQTVSLNMATGMD 597

Db    595 alvalgavdtalyavggwshkpldmskvfevidsynlgcgpgggddapqvfqdaglafsd 654
          : |||:|||:|:|:|    : |||:  :||: ::  :||||:||| : : ||: |||:|
Qy    598 SWVALAAVDSAVYGVQRGAKKPLE--RVFQFLEKSDLGCGAGGGLNNANVFHLAGLTFLT 655

Db    655 gdrltqtre-dlscpkekksrqkrnvnflkalseklgqyesspdakrccqdgmtklpmkrt 713
              :::| | :| ||   | :| :    | : |   : |  :   |:|| ||    :     |
Qy    656 NANADDSQENDEPC-KEIL-RPRRTLQ--KKIEEIAAKYKHSVVKKCCYDGAC-VNNDET 710

Db    714 ceqraarvpqqa-crepflscckfaedlrrnqtrsqahlarkvrdmvnlieeddilvrts 772
          |||||||::  : |  :|  ||  | :|| | :: : :|:| :: |  |:        :|:
Qy    711 CEQRAARISLGPRCIKAFTECCVVASQLRANISHKDMQLGR-LH-MKTLLPVSKPEIRSY 768

Db    773 fpdnwlwrvepvdsskllltvrlpdsmttweihgvslsksk glcvakptrvrvfrkfhlhl 832
          ||::||| |  |    | |   ||||:|||||:|:::|:: |:|||    : :||:   | :
Qy    769 FPESWLWEVHLVPRRKQLQFALPDSLTTWEIQGIGISNT-GICVADTVKAKVFKDVFLEM 827

Db    833 rlpisvrrfeqlelrpvlynylnddknvsvhvtpveglcmagggmmaqwvivpagsalp- 891
          :| || | ||::|: :|||    :     | :::|||:| : : :: :    ::
Qy    828 NIPYSVVRGEQIQLKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTKSSKCVRQK 887

Db    892 vafsvvptas-tnvplkl-vak-g-tlds--sdsv-skilqiekegaihreeiv-ynldp 943
          |  |     : | :|| : : :      :|:: :   :  | |:: || : ||     |||
Qy    888 VEGSSSHLVTFTVLPLEIGLHNINFSLETWFGKEILVKTLRVVPEG-VKRESYSGVTLDP 946

Db    944 --l-nnlgqmleipgssdpnivpegdfstfvkvtasepletlgsegalspggvasllrlp 1000
            :  ::: |:|      ::||  ::  :: | :            | | | : || |:  | :||
Qy    947 RGIYGTISRRKEFPYRIPLDLVPKTEIKRILSVKGLLVGEIL-S-AVLSQEGINILTHLP 1004

Db   1001 rscaeqtmiylapltasnyldrtkqwsklspet-kd-havhliqk-ghvriqqfrkkdg 1057
          :: ||  :: : | :    :||:  ::|: : ::    : : :      | |: | :|: |
Qy   1005 KGSAEAELMSVVPVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEGMLSIMSYRNADY 1064
```

```
*****************************************************************************
 MPSRCH (TM)
*****************************************************************************
```

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

```
MPsrch_pp   protein - protein database search, using Smith-Waterman algorithm

Run on:          Wed Jan 28 12:16:06 1998;  MasPar time 37.99 Seconds
                                            935.704 Million cell updates/sec
Tabular output not generated.

Title:           >US-08-487-283A-2
Description:     (1-1676) from US08487283A.pep
Perfect Score:   12048
Sequence:        1 MGLLGILCFLIFLGKTWGQE..........CQAFLANLDEFAEDIFLNGC 1676

Scoring table:   PAM 150
                 Gap 11

Searched:        59021 seqs, 21210388 residues

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database:        swiss-prot34
                   1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
                   8:part8 9:part9 10:part10 11:part11

Statistics:      Mean 58.177;  Variance 98.799;  scale 0.589
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	12048	100.0	1676	2	CO5_HUMAN	COMPLEMENT C5 PRECURS	0.00e+00
2	9789	81.3	1680	2	CO5_MOUSE	COMPLEMENT C5 PRECURS	0.00e+00
3	2367	19.6	1651	2	CO3_NAJNA	COMPLEMENT C3 PRECURS	0.00e+00
4	2361	19.6	1663	2	CO3_HUMAN	COMPLEMENT C3 PRECURS	0.00e+00
5	2309	19.2	1673	2	CO3_LAMJA	COMPLEMENT C3 PRECURS	0.00e+00
6	2296	19.1	1663	2	CO3_MOUSE	COMPLEMENT C3 PRECURS	0.00e+00
7	2302	19.1	1666	2	CO3_CAVPO	COMPLEMENT C3 PRECURS	0.00e+00
8	2255	18.7	1663	2	CO3_RAT	COMPLEMENT C3 PRECURS	0.00e+00
9	2106	17.5	1620	2	CO3_EPTBU	COMPLEMENT C3 (CONTAI	0.00e+00
10	2002	16.6	1640	2	CO3_ONCMY	COMPLEMENT C3-1 (CONT	0.00e+00
11	1723	14.3	1741	2	CO4_HUMAN	COMPLEMENT C4 PRECURS	0.00e+00
12	1605	13.3	1738	2	CO4_MOUSE	COMPLEMENT C4 PRECURS	0.00e+00
13	796	6.6	726	2	CO3_RABIT	COMPLEMENT C3 ALPHA C	5.45e-148

14	670	5.6	1477	1	A1I3_RAT	ALPHA-1-INHIBITOR III	2.99e-118
15	661	5.5	1476	1	A2MG_MOUSE	MURINOGLOBULIN 1 PREC	3.83e-116
16	592	4.8	1474	1	A2MG_HUMAN	ALPHA-2-MACROGLOBULIN	8.97e-98
17	581	4.8	1482	8	PZP_HUMAN	PREGNANCY ZONE PROTEI	1.53e-97
18	508	4.2	323	2	CO3_XENLA	COMPLEMENT C3 (FRAGME	8.31e-81
19	507	4.2	1472	1	A2MG_RAT	ALPHA-2-MACROGLOBULIN	1.40e-80
20	502	4.2	1473	7	OVOS_CHICK	OVOSTATIN PRECURSOR (	1.92e-79
21	493	4.1	1451	1	A2MH_MOUSE	MURINOGLOBULIN 2 PREC	2.11e-77
22	399	3.3	74	2	CO5A_PIG	COMPLEMENT C5A ANAPHY	2.09e-56
23	385	3.2	74	2	CO5A_BOVIN	COMPLEMENT C5A ANAPHY	2.44e-53
24	351	2.9	76	2	CO5A_RAT	COMPLEMENT C5A ANAPHY	5.68e-46
25	171	1.4	76	2	CO4A_RAT	COMPLEMENT C4A ANAPHY	3.12e-10
26	170	1.4	77	2	CO4A_BOVIN	COMPLEMENT C4A ANAPHY	4.64e-10
27	139	1.2	77	2	CO3A_PIG	COMPLEMENT C3A ANAPHY	5.44e-05
28	121	1.0	798	10	YA73_SCHPO	HYPOTHETICAL 92.1 KD	2.39e-02
29	115	1.0	1292	8	RPOC_MYCGE	DNA-DIRECTED RNA POLY	1.57e-01
30	105	0.9	257	11	YGBI_HAEIN	HYPOTHETICAL TRANSCRI	3.00e+00
31	106	0.9	270	11	YD1J_SCHPO	HYPOTHETICAL 31.5 KD	2.26e+00
32	105	0.9	328	1	ASCD_YERPS	CDP-6-DEOXY-DELTA-3,4	3.00e+00
33	110	0.9	341	11	YEJK_HAEIN	HYPOTHETICAL PROTEIN	7.10e-01
34	108	0.9	461	1	AD4B_BOVIN	STEROID HORMONE RECEP	1.27e+00
35	106	0.9	462	1	AD4B_MOUSE	STEROID HORMONE RECEP	2.26e+00
36	112	0.9	472	8	PSBC_SYNY3	PHOTOSYSTEM II 44 KD	3.91e-01
37	108	0.9	589	5	KPYK_LACDE	PYRUVATE KINASE (EC 2	1.27e+00
38	108	0.9	612	10	UNC6_CAEEL	UNC-6 PROTEIN PRECURS	1.27e+00
39	108	0.9	943	5	IROA_NEIME	IRON-REGULATED OUTER	1.27e+00
40	113	0.9	1024	10	UBA1_YEAST	UBIQUITIN-ACTIVATING	2.90e-01
41	106	0.9	1030	10	VPP1_CAEEL	PUTATIVE CLATHRIN-COA	2.26e+00
42	112	0.9	1517	4	GLTB_ECOLI	GLUTAMATE SYNTHASE (N	3.91e-01
43	107	0.9	1736	11	ZO1_HUMAN	TIGHT JUNCTION PROTEI	1.70e+00
44	108	0.9	1805	10	Y218_MYCGE	HYPOTHETICAL PROTEIN	1.27e+00
45	106	0:9	5147	3	FAT_DROME	CADHERIN-RELATED TUMO	2.26e+00

ALIGNMENTS

```
RESULT    1
ID   CO5_HUMAN      STANDARD;      PRT;  1676 AA.
AC   P01031;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   COMPLEMENT C5 PRECURSOR (CONTAINS: C5A ANAPHYLATOXIN).
GN   C5.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91079575.
RA   HAVILAND D.L., HAVILAND J.C., FLEISCHER D.T., HUNT A., WETSEL R.A.;
RL   J. IMMUNOL. 146:362-368(1991).
RN   [2]
RP   SEQUENCE OF 412-1676 FROM N.A.
RX   MEDLINE; 88209511.
RA   WETSEL R.A., LEMONS R.S., LEBEAU M.M., BARNUM S.R., NOACK D.,
RA   TACK B.F.;
RL   BIOCHEMISTRY 27:1474-1482(1988).
RN   [3]
RP   SEQUENCE OF 412-902 FROM N.A.
RX   MEDLINE; 85130937.
RA   LUNDWALL A.B., WETSEL R.A., KRISTENSEN T., WHITEHEAD A.S.,
RA   WOODS D.E., OGDEN R.C., COLTEN H.R., TACK B.F.;
RL   J. BIOL. CHEM. 260:2108-2112(1985).
```

```
RN   [4]
RP   SEQUENCE OF 678-751.
RX   MEDLINE; 79005687.
RA   FERNANDEZ H.N., HUGLI T.E.;
RL   J. BIOL. CHEM. 253:6955-6964(1978).
RN   [5]
RP   SEQUENCE OF 678-751 FROM N.A.
RX   MEDLINE; 91144547.
RA   BOHNSACK J.F., MOLLISON K.W., BUKO A.M., ASHWORTH J.C., HILL H.R.;
RL   BIOCHEM. J. 273:635-640(1991).
RN   [6]
RP   STRUCTURE BY NMR OF C5A.
RX   MEDLINE; 88309754.
RA   ZUIDERWEG E.R., MOLLISON K.W., HENKIN J., CARTER G.W.;
RL   BIOCHEMISTRY 27:3568-3580(1988).
RN   [7]
RP   STRUCTURE BY NMR OF C5A.
RX   MEDLINE; 89207527.
RA   ZUIDERWEG E.R., NETTESHEIM D.G., MOLLISON K.W., CARTER G.W.;
RL   BIOCHEMISTRY 28:172-185(1989).
RN   [8]
RP   STRUCTURE BY NMR OF C5A.
RX   MEDLINE; 89274164.
RA   ZUIDERWEG E.R., FESIK S.W.;
RL   BIOCHEMISTRY 28:2387-2391(1989).
CC   -!- FUNCTION: ACTIVATION OF C5 BY A C5 CONVERTASE INITIATES THE
CC       SPONTANEOUS ASSEMBLY OF THE LATE COMPLEMENT COMPONENTS, C5-C9,
CC       INTO THE MEMBRANE ATTACK COMPLEX. C5B HAS A TRANSIENT BINDING SITE
CC       FOR C6. THE C5B-C6 COMPLEX IS THE FOUNDATION UPON WHICH THE LYTIC
CC       COMPLEX IS ASSEMBLED.
CC   -!- SUBUNIT: C5 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 BASIC
CC       RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE
CC       BOND. C5 CONVERTASE ACTIVATES C5 BY CLEAVING THE ALPHA CHAIN,
CC       RELEASING C5A ANAPHYLATOXIN & GENERATING C5B (BETA CHAIN + ALPHA'
CC       CHAIN).
CC   -!- SIMILARITY: TO C3, C4 AND ALPHA-2-MACROGLOBULIN.
CC   -!- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C5,
CC       C5 ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT
CC       INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
CC       PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
CC       BASOPHILIC LEUKOCYTES. C5A ALSO STIMULATES THE LOCOMOTION OF
CC       POLYMORPHONUCLEAR LEUKOCYTES (CHEMOKINESIS) AND DIRECT THEIR
CC       MIGRATION TOWARD SITES OF INFLAMMATION (CHEMOTAXIS).
CC   -!- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 855
CC       ONWARD DUE TO THE PRESENCE OF AN ALU REPEAT.
CC   -!- SIMILARITY: CONTAINS ONE ANAPHYLATOXIN-LIKE DOMAIN.
DR   EMBL; M57729; G179983; -.
DR   EMBL; M65134; G179692; -.
DR   PIR; A40075; C5HU.
DR   PIR; S15121; S15121.
DR   HSSP; P01032; 1C5A.
DR   MIM; 120900; -.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN.
KW   COMPLEMENT PATHWAY; COMPLEMENT ALTERNATE PATHWAY; GLYCOPROTEIN;
KW   PLASMA; MEMBRANE ATTACK COMPLEX; CYTOLYSIS; INFLAMMATORY RESPONSE;
KW   SIGNAL; POLYMORPHISM.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19    673       COMPLEMENT C5 BETA CHAIN.
FT   PROPEP      674    677
FT   CHAIN       678   1676       COMPLEMENT C5 ALPHA CHAIN.
FT   PEPTIDE     678    751       C5A ANAPHYLATOXIN.
FT   CHAIN       752   1676       C5B (ALPHA').
FT   DOMAIN      698    732       ANAPHYLATOXIN-LIKE.
FT   DISULFID    698    724
```



```
FT   DISULFID    699    731
FT   DISULFID    711    732
FT   CARBOHYD    741    741
FT   CARBOHYD    911    911       POTENTIAL.
FT   CARBOHYD   1115   1115       POTENTIAL.
FT   CARBOHYD   1630   1630       POTENTIAL.
FT   VARIANT     518    518       F -> S.
SQ   SEQUENCE   1676 AA;  188331 MW;  9D5C6E59 CRC32;

  Query Match             100.0%;  Score 12048;  DB 2;  Length 1676;
  Best Local Similarity   100.0%;  Pred. No. 0.00e+00;
  Matches 1676;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Db      1 mgllgilcfliflgktwgqeqtyvisapkifrvgaseniviqvygyteafdatisiksyp 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 MGLLGILCFLIFLGKTWGQEQTYVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSYP 60

Db     61 dkkfsyssghvhlssenkfqnsailtiqpkqlpggqnpvsyvylevvskhfskskrmpit 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 DKKFSYSSGHVHLSSENKFQNSAILTIQPKQLPGGQNPVSYVYLEVVSKHFSKSKRMPIT 120

Db    121 ydngflfihtdkpvytpdqsvkvrvyslnddlkpakretvltfidpegsevdmveeidhi 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    121 YDNGFLFIHTDKPVYTPDQSVKVRVYSLNDDLKPAKRETVLTFIDPEGSEVDMVEEIDHI 180

Db    181 giisfpdfkipsnprygmwtikakykedfsttgtayfevkeyvlphfsvsiepeynfigy 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    181 GIISFPDFKIPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEPEYNFIGY 240

Db    241 knfknfeitikaryfynkvvteadvyitfgiredlkddqkemmqtamqntmlingiaqvt 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    241 KNFKNFEITIKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLINGIAQVT 300

Db    301 fdsetavkelsyysledlnnkylyiavtviestggfseeaeipgikyvlspyklnlvatp 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    301 FDSETAVKELSYYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLNLVATP 360

Db    361 lflkpgipypikvqvkdsldqlvggvpvilnaqtidvnqetsdldpsksvtrvddgvasf 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    361 LFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDDGVASF 420

Db    421 vlnlpsgvtvlefnvktdapdlpeenqaregyraiaysslsqsylyidwtdnhkallvge 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    421 VLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKALLVGE 480

Db    481 hlniivtpkspyidkithynylilskgkiihfgtrekfsdasyqsinipvtqnmvpssrl 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    481 HLNIIVTPKSPYIDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQNMVPSSRL 540

Db    541 lvyyivtgeqtaelvsdsvwlnieekcgnqlqvhlspdadayspgqtvslnmatgmdswv 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    541 LVYYIVTGEQTAELVSDSVWLNIEEKCGNQLQVHLSPDADAYSPGQTVSLNMATGMDSWV 600

Db    601 alaavdsavygvqrgakkplervfqflekedlgcgagggl nnanvfhlagltfltnanad 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    601 ALAAVDSAVYGVQRGAKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLAGLTFLTNANAD 660

Db    661 dsqendepckeilrprrtlqkkieeiaakykhsvvkkccydgacvnndetceqraarisl 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    661 DSQENDEPCKEILRPRRTLQKKIEEIAAKYKHSVVKKCCYDGACVNNDETCEQRAARISL 720

Db    721 gprcikafteccvvasqlranishkdmqlgrlhmktllpvskpeirsyfpeswlwevhlv 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    721 GPRCIKAFTECCVVASQLRANISHKDMQLGRLHMKTLLPVSKPEIRSYFPESWLWEVHLV 780
```

```
Db    781 prrkqlqfalpdslttweiqgigisntgicvadtvkakvfkdvflemnipysvvrgeqiq 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    781 PRRKQLQFALPDSLTTWEIQGIGISNTGICVADTVKAKVFKDVFLEMNIPYSVVRGEQIQ 840

Db    841 lkgtvynyrtsgmqfcvkmsavegictsespvidhqgtksskcvrqkvegsssshlvtftv 900
          |||||||||||||||||||||||||||||||||||||,||||||||||||||||||||||
Qy    841 LKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTKSSKCVRQKVEGSSSHLVTFTV 900

Db    901 lpleiglhninfsletwfgkeilvktlrvvpegvkresysgvtldprgiygtisrrkefp 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    901 LPLEIGLHNINFSLETWFGKEILVKTLRVVPEGVKRESYSGVTLDPRGIYGTISRRKEFP 960

Db    961 yripldlvpkteikrilsvkgllvgeilsavlsqeginilthlpkgsaeaelmsvvpvfy 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    961 YRIPLDLVPKTEIKRILSVKGLLVGEILSAVLSQEGINILTHLPKGSAEAELMSVVPVFY 1020

Db   1021 vfhyletgnhwnifhsdpliekqklkkklkegmlsimsyrnadysysvwkggsastwlta 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1021 VFHYLETGNHWNIFHSDPLIEKQKLKKKLKEGMLSIMSYRNADYSYSVWKGGSASTWLTA 1080

Db   1081 falrvlgqvnkyveqnqnsicnsllwlvenyqldngsfkensqyqpiklqgtlpvearen 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1081 FALRVLGQVNKYVEQNQNSICNSLLWLVENYQLDNGSFKENSQYQPIKLQGTLPVEAREN 1140

Db   1141 slyltaftvigirkafdicplvkidtalikadnfllentlpaqstftlaisayalslgdk 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1141 SLYLTAFTVIGIRKAFDICPLVKIDTALIKADNFLLENTLPAQSTFTLAISAYALSLGDK 1200

Db   1201 thpqfrsivsalkrealvkgnppiyrfwkdnlqhkdssvpntgtarmvettayalltsln 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1201 THPQFRSIVSALKREALVKGNPPIYRFWKDNLQHKDSSVPNTGTARMVETTAYALLTSLN 1260

Db   1261 lkdinyvnpvikwlseeqryggg fystqdtinaieglteysllvkqlrlsmdidvsykhk 1320
          |||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Qy   1261 LKDINYVNPVIKWLSEEQRYGGGFYSTQDTINAIEGLTEYSLLVKQLRLSMDIDVSYKHK 1320

Db   1321 galhnykmtdknflgrpvevllnddlivstgfgsglatvhvttvvhktstseevcsfylk 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1321 GALHNYKMTDKNFLGRPVEVLLNDDLIVSTGFGSGLATVHVTTVVHKTSTSEEVCSFYLK 1380

Db   1381 idtqdieashyrgygnsdykrivacasykpsreesssgsshavmdislptgisaneedlk 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1381 IDTQDIEASHYRGYGNSDYKRIVACASYKPSREESSSGSSHAVMDISLPTGISANEEDLK 1440

Db   1441 alvegvdqlftdyqikdghvilqlnsipssdflcvrfrifelfevgflspatftvyeyhr 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1441 ALVEGVDQLFTDYQIKDGHVILQLNSIPSSDFLCVRFRIFELFEVGFLSPATFTVYEYHR 1500

Db   1501 pdkqctmfystsnikiqkvcegaackcveadcgqmqeeldltisaetrkqtackpeiaya 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1501 PDKQCTMFYSTSNIKIQKVCEGAACKCVEADCGQMQEELDLTISAETRKQTACKPEIAYA 1560

Db   1561 ykvsitsitvenvfvkykatlldiyktgeavaekdseitfikkvtctnaelvkgrqylim 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1561 YKVSITSITVENVFVKYKATLLDIYKTGEAVAEKDSEITFIKKVTCTNAELVKGRQYLIM 1620

Db   1621 gkealqikynfsfryiypldsltwieywprdttcsscqaflanldefaedifIngc 1676
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1621 GKEALQIKYNFSFRYIYPLDSLTWIEYWPRDTTCSSCQAFLANLDEFAEDIFLNGC 1676


RESULT    2
ID   CO5_MOUSE      STANDARD;      PRT;  1680 AA.
```



```
AC   P06684;
DT   01-JAN-1988 (REL. 06, CREATED)
DT   01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   COMPLEMENT C5 PRECURSOR (CONTAINS: C5A ANAPHYLATOXIN).
GN   C5.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90153853.
RA   WETSEL R.A., FLEISCHER D.T., HAVILAND D.L.;
RL   J. BIOL. CHEM. 265:2435-2440(1990).
RN   [2]
RP   SEQUENCE OF 41-1680 FROM N.A.
RX   MEDLINE; 87185363.
RA   WETSEL R.A., OGATA R.T., TACK B.F.;
RL   BIOCHEMISTRY 26:737-743(1987).
CC   -!- FUNCTION: ACTIVATION OF C5 BY A C5 CONVERTASE INITIATES THE
CC       SPONTANEOUS ASSEMBLY OF THE LATE COMPLEMENT COMPONENTS, C5-C9,
CC       INTO THE MEMBRANE ATTACK COMPLEX. C5B HAS A TRANSIENT BINDING SITE
CC       FOR C6. THE C5B-C6 COMPLEX IS THE FOUNDATION UPON WHICH THE LYTIC
CC       COMPLEX IS ASSEMBLED.
CC   -!- SUBUNIT: C5 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 BASIC
CC       RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE
CC       BOND. C5 CONVERTASE ACTIVATES C5 BY CLEAVING THE ALPHA CHAIN,
CC       RELEASING C5A ANAPHYLATOXIN & GENERATING C5B (BETA CHAIN + ALPHA'
CC       CHAIN).
CC   -!- SIMILARITY: TO C3, C4 AND ALPHA-2-MACROGLOBULIN.
CC   -!- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C5,
CC       C5 ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT
CC       INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
CC       PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
CC       BASOPHILIC LEUKOCYTES. C5A ALSO STIMULATES THE LOCOMOTION OF
CC       POLYMORPHONUCLEAR LEUKOCYTES (CHEMOKINESIS) AND DIRECT THEIR
CC       MIGRATION TOWARD SITES OF INFLAMMATION (CHEMOTAXIS).
CC   -!- SIMILARITY: CONTAINS ONE ANAPHYLATOXIN-LIKE DOMAIN.
DR   EMBL; M35525; G309124; -.
DR   EMBL; M35526; G309123; -.
DR   PIR; A27538; A27538.
DR   PIR; A35530; A35530.
DR   HSSP; P01032; 1C5A.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN.
KW   COMPLEMENT PATHWAY; COMPLEMENT ALTERNATE PATHWAY; GLYCOPROTEIN;
KW   PLASMA; MEMBRANE ATTACK COMPLEX; CYTOLYSIS; INFLAMMATORY RESPONSE;
KW   SIGNAL.
FT   SIGNAL        1     18
FT   CHAIN        19   1680       COMPLEMENT C5.
FT   CHAIN        19    674       COMPLEMENT C5 BETA CHAIN.
FT   PROPEP      675    678
FT   CHAIN       679   1680       COMPLEMENT C5 ALPHA CHAIN.
FT   PEPTIDE     679    755       C5A ANAPHYLATOXIN.
FT   CHAIN       756   1680       C5B (ALPHA').
FT   DOMAIN      702    736       ANAPHYLATOXIN-LIKE.
FT   DISULFID    702    728       BY SIMILARITY.
FT   DISULFID    703    735       BY SIMILARITY.
FT   DISULFID    715    736       BY SIMILARITY.
FT   CARBOHYD    427    427       POTENTIAL.
FT   CARBOHYD    915    915       POTENTIAL.
FT   CARBOHYD   1119   1119       POTENTIAL.
FT   CARBOHYD   1633   1633       POTENTIAL.
FT   VARIANT     216    216       Y -> L (IN DEFECTIVE VARIANT C5D).
FT   VARIANT     217   1680       MISSING (IN DEFECTIVE VARIANT C5D).
```

```
SQ   SEQUENCE   1680 AA;  188877 MW;  AA17044B CRC32;

  Query Match             81.3%;  Score 9789;  DB 2;  Length 1680;
  Best Local Similarity   77.8%;  Pred. No. 0.00e+00;
  Matches 1307;  Conservative  218;  Mismatches 149;  Indels    7;  Gaps    6;

Db          1 mglwgilcllifldktwgqeqtyvisapkilrvgssenvviqvhgyteafdatlslksyp 60
              ||| ||||:|||| |||||||||||||||:|||:|||:|||| |||||||||:|:||||
Qy          1 MGLLGILCFLIFLGKTWGQEQTYVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSYP 60

Db         61 dkkvtfssgyvnlspenkfqnaalltlqpnqvpreespvshvylevvskhfskskkipit 120
              ||| ::||| |:||:||||||:|:||:||:|:|   ::||| |||||||||||||::|||
Qy         61 DKKFSYSSGHVHLSSENKFQNSAILTIQPKQLPGGQNPVSYVYLEVVSKHFSKSKRMPIT 120

Db        121 ynngilfihtdkpvytpdqsvkirvyslgddlkpakretvltfidpegsevdiveendyt 180
              |:||:||||||||||||||||||:|||| |||||||||||||||||||||||:||| |
Qy        121 YDNGFLFIHTDKPVYTPDQSVKVRVYSLNDDLKPAKRETVLTFIDPEGSEVDMVEEIDHI 180

Db        181 giisfpdfkipsnpkygvwtikanykkdfttgtayfeikeyvlprfsvsielertfigy 240
              ||||||||||||||:||:|||||:|| ||:||||||||:|||||||:|||||| |  ||||
Qy        181 GIISFPDFKIPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEPEYNFIGY 240

Db        241 knfknfeitvkaryfynkvvpdaevyaffglredikdeekqmmhkatqaaklvdgvaqis 300
              ||||||||:|||||||||| :|:|| ||:|||:||::|:||: | | ::|::|:||::
Qy        241 KNFKNFEITIKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLINGIAQVT 300

Db        301 fdsetavkelsynsledlnnkylyiavtvtessggfseeaeipgvkyvlspytlnlvatp 360
              ||||||||||| |||||||||||||||| ||:|||||||||||:|||||||| |||||||
Qy        301 FDSETAVKELSYYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLNLVATP 360

Db        361 lfvkpgipfsikaqvkdsleqavggvpvtlmaqtvdvnqetsdletkrsithdtdgvavf 420
              ||:|||||::|| ||||||:| |||||| | |||:||||||||:  :|:|:  |||| |
Qy        361 LFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDDGVASF 420

Db        421 vlnlpsnvtvlkfeirtddpelpeenqaskeyeavayssslsqsyiyiawtenykpmlvge 480
              |||||| |||| |:::|| |:|||||||    | |:||||||||||:|| ||:| |::||||
Qy        421 VLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKALLVGE 480

Db        481 ylnimvtpkspyidkithynylilskgkivqygtreklfsstyqninipvtqnmvpsarl 540
              |||:||||||||||||||||||||||||:::||||||:  ::||:|||||||||||:||
Qy        481 HLNIIVTPKSPYIDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQNMVPSSRL 540

Db        541 lvyyivtgeqtaelvadavwinieekcgnqlqvhlspdeyvyspgqtvsldmvteadswv 600
              ||||||||||||||:|:||:||||||||||||||||    |||||||||:| |  ||||
Qy        541 LVYYIVTGEQTAELVSDSVWLNIEEKCGNQLQVHLSPDADAYSPGQTVSLNMATGMDSWV 600

Db        601 alsavdravykvqgnakramqrvfqaldekedlgcgaggghdnadvfhlagltfltnana 660
              ||:|||  |||.||  ||::::||||  | |||||||||:|| :||:|||||||||||||||
Qy        601 ALAAVDSAVYGVQRGAKKPLERVFQFL-EKSDLGCGAGGGLNNANVFHLAGLTFLTNANA 659

Db        661 ddshyrddsckeilrskrnlhllrqkieeqaakykhsvpkkccydgarvnfyetceerva 720
              |||:  |::||||||::| |:  : |||| ||||||| ||||||| ||  ||||:| |
Qy        660 DDSQENDEPCKEILRPRRTLQ--K-KIEEIAAKYKHSVVKKCCYDGACVNNDETCEQRAA 716

Db        721 rvtigplcirafnecctiankirkesphkpvqlgrihiktllpvmkadirsyfpeswlwe 780
              |:::|| ||:|| |||  :|: :| : : :|| :||||:|:|||||| |::|||||||||||
Qy        717 RISLGPRCIKAFTECCVVASQLRANISHKDMQLGRLHMKTLLPVSKPEIRSYFPESWLWE 776

Db        781 ihrvpkrkqlqvtlpdslttweiqgigisdngicvadtlkakvfkevflemnipysvvrg 840
              :| ||:||||| :||||||||||||||||: ||||||:||||||:||||||||||||||
Qy        777 VHLVPRRKQLQFALPDSLTTWEIQGIGISNTGICVADTVKAKVFKDVFLEMNIPYSVVRG 836

Db        841 eqiqlkgtvynymtsgtkfcvkmsavegictsgssaaslhtsrpsrcvfqriegssshlv 900
              |||||||||| ||| |||||||||||||| |:     : :::|:|| |::||||||||
```



```
Qy        837 EQIQLKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTKSSKCVRQKVEGSSSHLV 896

Db        901 tftllpleiglhsinfsletsfgkdilvktlrvvpegvkresyagvildpkgirgivnrr 960
              |||:||||||||:||||||| |||:||||||||||||||||||:|| |||:|| | ::||
Qy        897 TFTVLPLEIGLHNINFSLETWFGKEILVKTLRVVPEGVKRESYSGVTLDPRGIYGTISRR 956

Db        961 kefpyripldlvpktkverilsvkgllvgeflstvlskeginilthlpkgsaeaelmsia 1020
              |||||||||||||| : |||||||||||:||:||| ||||||||||||||||||:
Qy        957 KEFPYRIPLDLVPKTEIKRILSVKGLLVGEILSAVLSQEGINILTHLPKGSAEAELMSVV 1016

Db       1021 pvfyvfhyleagnhwnifypdtlskrqslekkikqgvvsvmsyrnadysysmwkgasast 1080
              |||||||||:||||||| :| |   :| | ||:|:|::|:|||||||||||||:|||:||||
Qy       1017 PVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEGMLSIMSYRNADYSYSVWKGGSAST 1076

Db       1081 wltafalrvlgqvakyvkqdensicnsllwlvekcqlengsfkensqylpiklqgtlpae 1140
              |||||||||||| ||| |::||||||||||||: ||:||||||||| |||||||| |
Qy       1077 WLTAFALRVLGQVNKYVEQNQNSICNSLLWLVENYQLDNGSFKENSQYQPIKLQGTLPVE 1136

Db       1141 aqektlylytafsvigirkavdicptmkihtaldkadsfllentlpskstftlaivayals 1200
              |:|::|||||:|||||| |||| :|| ||| |||:||||||||: |||||| |||||
Qy       1137 ARENSLYLTAFTVIGIRKAFDICPLVKIDTALIKADNFLLENTLPAQSTFTLAISAYALS 1196

Db       1201 lgdrthprfrlivsalrkeafvkgdppiyrywrdtlkrpdssvpssgtagmvettayall 1260
              |||:|||:|| |||||::|:|||:||||:|:| | : |||||::||| |||||||||
Qy       1197 LGDKTHPQFRSIVSALKREALVKGNPPIYRFWKDNLQHKDSSVPNTGTARMVETTAYALL 1256

Db       1261 aslklkdmnyanpiikwlseeqryggGfystqdtinaiegltey slllkqihldmdinva 1320
              :||:|||:|| |||:||||||||||||||||||||||||||||||||:|||::| |||:|:
Qy       1257 TSLNLKDINYVNPVIKWLSEEQRYGGGFYSTQDTINAIEGLTEYSLLVKQLRLSMDIDVS 1316

Db       1321 ykhegdfhkykvtekhflgrpvevslnddlvvstgyssglatvyvktvvhkisvseefcs 1380
              ||| | :|:||:|:|:||||||| ||||:|||:||||| | ||||| | ||| ||
Qy       1317 YKHKGALHNYKMTDKNFLGRPVEVLLNDDLIVSTGFGSGLATVHVTTVVHKTSTSEEVCS 1376

Db       1381 fylkidtqdieasshfr-lsdsgfkriiacasykpskeestsgsshavmdislptgigan 1439
              |||||||||||| |:|  ::| :|||:|||||||||:|||:|||||||||||||||:||
Qy       1377 FYLKIDTQDIEAS-HYRGYGNSDYKRIVACASYKPSREESSSGSSHAVMDISLPTGISAN 1435

Db       1440 eedlralvegvdqlltdyqikdghvilqlnsipsrdflcvrfrifelfqvgflnpatftv 1499
              ||||:||||||||:||||||||||||||||||||| ||||||||||||:||||:||||||
Qy       1436 EEDLKALVEGVDQLFTDYQIKDGHVILQLNSIPSSDFLCVRFRIFELFEVGFLSPATFTV 1495

Db       1500 yeyhrpdkqctmiysisdtrlqkvcegaactcveadcaqlqaevdlaisadsrkekackp 1559
              ||||||||||||:|| |: ::||||||||| ||||||:|:| |:||:|||::||: ||||
Qy       1496 YEYHRPDKQCTMFYSTSNIKIQKVCEGAACKCVEADCGQMQEELDLTISAETRKQTACKP 1555

Db       1560 etayaykvritsateenvfvkytatllvtyktgea-adensevtfikkmsctnanlvkgk 1618
              | |||||| ||| | |||||| ||||  ||||||  |: :||:||||::||||:||||:
Qy       1556 EIAYAYKVSITSITVENVFVKYKATLLDIYKTGEAVAEKDSEITFIKKVTCTNAELVKGR 1615

Db       1619 qylimgkevlqikhnfsfkyiypldsstwieywptdttcpscqafvenlnnfaedlflns 1678
              ||||||| |||| ||||:|||||||| |||||| ||||:|||||: ||::||||:|||:
Qy       1616 QYLIMGKEALQIKYNFSFRYIYPLDSLTWIEYWPRDTTCSSCQAFLANLDEFAEDIFLNG 1675

Db       1679 c 1679
              |
Qy       1676 C 1676

RESULT    3
ID   CO3_NAJNA      STANDARD;      PRT;  1651 AA.
AC   Q01833;
DT   01-JUL-1993 (REL. 26, CREATED)
DT   01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
```

```
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   COMPLEMENT C3 PRECURSOR (CONTAINS: C3A ANAPHYLATOXIN).
GN   C3.
OS   NAJA NAJA (INDIAN COBRA).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; REPTILIA;
OC   LEPIDOSAURIA; SERPENTES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 93056528.
RA   FRITZINGER D.C., CONNELLY M., PETRELLA E.C., BREDEHORST R.,
RA   VOGEL C.W.;
RL   J. IMMUNOL. 149:3554-3562(1992).
CC   -!- FUNCTION: C3 PLAYS A CENTRAL ROLE IN THE ACTIVATION OF THE
CC       COMPLEMENT SYSTEM. ITS PROCESSING BY C3 CONVERTASE IS THE CENTRAL
CC       REACTION IN BOTH CLASSICAL AND ALTERNATIVE COMPLEMENT PATHWAYS.
CC       AFTER ACTIVATION C3B CAN BIND COVALENTLY, VIA ITS REACTIVE
CC       THIOLESTER, TO CELL SURFACE CARBOHYDRATES OR IMMUNE AGGREGATES.
CC   -!- SUBUNIT: C3 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 ARG
CC       RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE
CC       BOND. C3 CONVERTASE ACTIVATES C3 BY CLEAVING THE ALPHA CHAIN,
CC       RELEASING C3A ANAPHYLATOXIN & GENERATING C3B (BETA CHAIN + ALPHA'
CC       CHAIN).
CC   -!- SIMILARITY: TO C4, C5 AND ALPHA-2-MACROGLOBULIN.
CC   -!- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C3,
CC       C3A ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT
CC       INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
CC       PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
CC       BASOPHILIC LEUKOCYTES.
CC   -!- SIMILARITY: CONTAINS ONE ANAPHYLATOXIN-LIKE DOMAIN.
DR   EMBL; L02365; G213373; -.
DR   PIR; A46513; A46513.
DR   HSSP; P01032; 1C5A.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN.
KW   COMPLEMENT PATHWAY; COMPLEMENT ALTERNATE PATHWAY; PLASMA;
KW   INFLAMMATORY RESPONSE; SIGNAL.
FT   SIGNAL        1     22
FT   CHAIN        23   1651       COMPLEMENT C3.
FT   CHAIN        23    655       BETA CHAIN.
FT   CHAIN       661   1651       ALPHA CHAIN.
FT   PEPTIDE     661    738       C3A ANAPHYLATOXIN.
FT   CHAIN       739   1651       C3B (ALPHA' CHAIN).
FT   SITE        738    739       CLEAVAGE (BY C3 CONVERTASE).
FT   DOMAIN      683    718       ANAPHYLATOXIN-LIKE.
FT   DISULFID    546    807       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    615    650       BY SIMILARITY.
FT   DISULFID    683    710       BY SIMILARITY.
FT   DISULFID    684    717       BY SIMILARITY.
FT   DISULFID    697    718       BY SIMILARITY.
FT   DISULFID    863   1501       BY SIMILARITY.
FT   DISULFID   1091   1147       BY SIMILARITY.
FT   DISULFID   1346   1477       BY SIMILARITY.
FT   DISULFID   1377   1446       BY SIMILARITY.
FT   DISULFID   1494   1499       BY SIMILARITY.
FT   DISULFID   1506   1578       BY SIMILARITY.
FT   DISULFID   1525   1649       BY SIMILARITY.
FT   THIOLEST    999   1002       BY SIMILARITY.
SQ   SEQUENCE   1651 AA;  184926 MW;  2958575F CRC32;

  Query Match            19.6%;  Score 2367;  DB 2;  Length 1651;
  Best Local Similarity  30.2%;  Pred. No. 0.00e+00;
  Matches  518;  Conservative  416;  Mismatches  655;  Indels  127;  Gaps  97;

Db      4 malylvaalligfpg-s-shgalytlitpavlrtdteeqilveahgdstpksldifvhdf 61
```



```
Qy      1 MGLLGILCFLI-FLGKTWGQEQTYVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSY 59

Db     62 prkqktlfqsrvdmnqagsmfvtpti-kvpakelnkdskqnqyvvvkvtgpqvalekvvl 120

Qy     60 PDKKFSYSSGHVHLS-SENKFQNSAILTIQPKQLPGGQNPVSYVYLEVVSKHFSKSKRMP 118

Db    121 lsyqsgfvfiqtdkgiytpgspvryrvfsvdhnmhrmdktvivefqtpegivvs-skpvn 179

Qy    119 ITYDNGFLFIHTDKPVYTPDQSVKVRVYSLNDDLKPAKRETVLTFIDPEGSEVDMVEEID 178

Db    180 psgsir-p-ynlpelvsfgtwkavakyehspeesytayfdvreyvlpsfevrlqpsdkfl 237

Qy    179 HIGIISFPDFKIPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEPEYNFI 238

Db    238 -yidgnknfhvsitarylygkkv-egvafvvfgvk--i-ddakksipdsltripiidgdg 292

Qy    239 GYKN-FKNFEITIKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLINGIA 297

Db    293 eatlkrdt-lrs-rfqdlnqlvghtlyvsvtvitesgsdmvvteqggihivtspyqiyft 350

Qy    298 QVTFDSETAVKELSYYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLNLV 357

Db    351 ktpkyfkpgmpyeltvyvtnp-d---gs-paa-hvp-v-vs-ea--ih-segtt-lsdgt 397

Qy    358 ATPLFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDDGV 417

Db    398 aklilntplniqslpitvrtnhgdlprerqaiksmtatayqtqggsenylhvaitsteik 457

Qy    418 ASFVLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKALL 477

Db    458 pgdnlpvnfnvrgnanslnqikyftylilnkgkifkvgrqprrdgqnlvtmnlhitpdli 517

Qy    478 VGEHLNI-I-VTPKSPYIDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQNMV 535

Db    518 psfrfvayyqv-gnn--eivadsvwvdvkdtcmgtlvvkgassrddriqkpgaamkikle 574

Qy    536 PSSRLLVYYIVTGEQTAELVSDSVWLNIEEKCGNQLQVH-LSPDADAY-SPGQTVSLNMA 593

Db    575 gdpgarvglvavdkavyvlndkykisqakiwdtieksdfgctagsgqnnlgvfedaglal 634

Qy    594 TGMDSWVALAAVDSAVYGVQRGAKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLAGLTF 653

Db    635 ttstnlntkqrsaakcpqpanrrrrssvllldskaskaaqfqdqglrkccedgmhenpmg 694

Qy    654 LTNANADDSQENDEPCKE-ILRPRRT---LQKKIEEIAAKYKHSVVKKCCYDGACVNNDE 709

Db    695 ytcekrakyiqegdackaaflecchyikgirdenqreselflarsdfedelfgddniisr 754

Qy    710 -TCEQRAARISLGPRCIKAFTECCVVASQLRA-NISHKDMQLGRLHMKTLLPVSKPEI-R 766

Db    755 sdfpeswlwlteeltgepnnqgissktvpfylrdsittwellavglsptkgicvaepyei 814

Qy    767 SYFPESWLWEVH-LV--PR-R----KQLQFALPDSLTTWEIQGIGISNT-GICVADTVKA 817

Db    815 tvmkdffidlrlpysvvkneqveirailynyadedi-y-vr---veliynpafcsasteg 869

Qy    818 KVFKDVFLEMNIPYSVVRGEQIQLKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQG 877

Db    870 qry-r-qqfpikalssravpfvivpleqglhdveviasvrgelasdgvrkklkvvpeger 927

Qy    878 TKSSKCVRQKVEGSSSHLVTFTVLPLEIGLHNINFSLETW--FGKEILVKTLRVVPEGVK 935

Db    928 knivtiieldpsvkgv-ggtqeltviankld-dkvpdtevetrisvlgdpvaqiiensid 985
```

```
                 ::  : : |||  :|: |          :: ::  | || ||:   :|| |  |::|:   :
Qy     936 RESYSGVTLDP--RGIYGTISRRKEFPYRIPLDLVPKTEIKRILSVKGLLVGEILSAVLS 993

Db     986 gsklnhliitpsgcgeqnmitmtpsviaty-yldatgqwenlgvdrrte-a-ikqimt-g 1041
                 :| |    | | :| ::::: | |:   : ||::  :|: :   |    :|   :   |
Qy     994 QEGINILTHLPKGSAEAELMSVVP-VFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEG 1052

Db    1042 yaqqmvykkadhsyaaftnrassswltayvvkvlamasnmvkdisheiicggvkwlilnr 1101
                 | |::|| ||:      ::|:||||: ::||:  :: | : ::: || :: ||: |
Qy    1053 MLSIMSYRNADYSYSVWKGGSASTWLTAFALRVLGQVNKYV-EQNQNSICNSLLWLVENY 1111

Db    1102 qqpdgvfkenapvihgemlggtkgaep-easl--tafivtallesrsvckeqinildssi 1158
               |  :| ||||:    :  :| ||    |: | ||   ||| | ::   :   :|     ::| |:::
Qy    1112 QLDNGSFKENSQ-YQPIKLQGTLPVEARENSLYLTAFTVIGIRKAFDIC-PLVKI-DTAL 1168

Db    1159 nkatdyllkkyeklqrpyttaltayalaaadrlndd-r--v--lm-aa-stgrn---r-w 1207
               || ::|| :     |  :| |::||||: :|: : : |   |  |:   |    |      | |
Qy    1169 IKADNFLLENTLPAQSTFTLAISAYALSLGDKTHPQFnSIVSALKREALVKGNPPIYRFW 1228

Db    1208 -e--ey-narthn------iegtsyallallkmkkfaevgpvvrwlidqkyyggtygqtq 1257
               :  :  ::    |        :| |:||||: |::| :   | ||::|| ::   ||| :   ||
Qy    1229 KDNLQHKDSSVPNTGTARMVETTAYALLTSLNLKDINYVNPVIKWLSEEQRYGGGFYSTQ 1288

Db    1258 atvmvfqalaeyeiqmpthqdlnldisiklperevperysindrnavqartvetklnedf 1317
               |:  :::|:|| : :  :  |::|| :      :      | : |:| : :| ||  ||:|:
Qy    1289 DTINAIEGLTEYSL-LVKQLRLSMDIDVSYKHKGALHNYKMTDKNFL-GRPVEVLLNDDL 1346

Db    1318 tvsas-gdgkatmtiltvynaqlredanvcnkfhldvsvenvelnlkqakggkaalrlki 1376
                  ||:: | | ||: : || :        :||: | | :   :::| :   :: |       |:
Qy    1347 IVSTGFGSGLATVHVTTVVH-KTSTSEEVCS-FYLKIDTQDIEASHYRGYGNSDYKRIVA 1404

Db    1377 ctry-lg--evds-tm-tiidismltgffpdaedlkrlsngvdryiskfeidnnmaqkgt 1431
               |: |   :   |   | :   :::|||: ||: :: |||| | :|||: :: ::| :: ::
Qy    1405 CASYKPSREESSSGSSHAVMDISLPTGISANEEDLKALVEGVDQLFTDYQI-KD-GH--- 1459

Db    1432 vviyldkvshsedeclhfkihkhfevgfiqpgsvkvysyynldeqctkfyhpdketglln 1491
               |:: |: :: |:  |::|:|    |||||: |::  ||  |   | |||:||   :   :
Qy    1460 VILQLNSIPSSDFLCVRFRIFELFEVGFLSPATFTVYEYHRPDKQCTMFYSTSNIK--IQ 1517

Db    1492 kichgnicrcaeetcsllnqq-k-kidlqlriqkacaqnvdyvyktkllrieekdgndiy 1549
               |:| |   |:| |  |: : ::      |   : | | ||   :: | ||   :   |     :     |
Qy    1518 KVCEGAACKCVEADCGQMQEELDLTISAETRKQTACKPEIAYAYKVSITSITVENVFVKY 1577

Db    1550 fmdvlevikggtdrnaqakarqyvsqrkcqealnlkldndyliwglssdlwpmk-d-dis 1607
                  :|:: | |              ::      | :| :|     :||| |       |  :| :   :
Qy    1578 KATLLDIYKTGEAVAEKDSEITFIKKVTCTNA-ELVKCRQYLIMG-KEAL-QIKYNFSFR 1634

Db    1608 ylit-kn-twierwpnedecqd-eef-qnlcddfaq 1639
               |:     : |||| || :   |    : |   || |:||:
Qy    1635 YIYPLDSLTWIEYWPRDTTCSSCQAFLANL-DEFAE 1669
```

```
RESULT    4
ID   CO3_HUMAN      STANDARD;      PRT;   1663 AA.
AC   P01024;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   COMPLEMENT C3 PRECURSOR (CONTAINS: C3A ANAPHYLATOXIN).
GN   C3.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
```



```
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85140166.
RA   DE BRUIJN M.H.L., FEY G.H.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 82:708-712(1985).
RN   [2]
RP   SEQUENCE OF 672-748.
RX   MEDLINE; 76069169.
RA   HUGLI T.E.;
RL   J. BIOL. CHEM. 250:8293-8301(1975).
RN   [3]
RP   SEQUENCE OF 1409-1563.
RX   MEDLINE; 88154452.
RA   DAOUDAKI M.E., BECHERER J.D., LAMBRIS J.D.;
RL   J. IMMUNOL. 140:1577-1580(1988).
RN   [4]
RP   SEQUENCE OF 988-1036.
RX   MEDLINE; 82174534.
RA   THOMAS M.L., JANATOVA J., GRAY W.R., TACK B.F.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 79:1054-1058(1982).
RN   [5]
RP   STRUCTURE BY NMR OF C3A.
RX   MEDLINE; 88276894.
RA   NETTESHEIM D.G., EDALJI R.P., MOLLISON K.W., GREER J., ZUIDERWEG E.R.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 85:5036-5040(1988).
RN   [6]
RP   MUTAGENESIS OF THIOESTER BOND REGION.
RX   MEDLINE; 92250565.
RA   ISAAC L., ISENMAN D.E.;
RL   J. BIOL. CHEM. 267:10062-10069(1992).
RN   [7]
RP   DISULFIDE BONDS.
RX   MEDLINE; 93106233.
RA   DOLMER K., SOTTRUP-JENSEN L.;
RL   FEBS LETT. 315:85-90(1993).
RN   [8]
RP   VARIANT C3F/S.
RX   MEDLINE; 89309808.
RA   POZNANSKY M.C., CLISSOLD P.M., LACHMANN P.J.;
RL   J. IMMUNOL. 143:1254-1258(1989).
RN   [9]
RP   ERRATUM (RETRACTION OF ABOVE ARTICLE).
RX   MEDLINE; 90063087.
RA   POZNANSKY M.C., CLISSOLD P.M., LACHMANN P.J.;
RL   J. IMMUNOL. 143:3860-3862(1989).
RN   [10]
RP   VARIANTS GLY-102 AND PRO-314.
RX   MEDLINE; 91011240.
RA   BOTTO M., YONG FONG K., SO A.K., KOCH C., WALPORT M.J.;
RL   J. EXP. MED. 172:1011-1017(1990).
RN   [11]
RP   VARIANT ASN-549.
RX   MEDLINE; 95050640.
RA   SINGER L., WHITEHEAD W.T., AKAMA H., KATZ Y., FISHELSON Z.,
RA   WETSEL R.A.;
RL   J. BIOL. CHEM. 269:28494-28499(1994).
RN   [12]
RP   VARIANT GLN-1320.
RA   WATANABE Y., MATSUI N., YAN K., NISHIMUKAI H., TOKUNAGA K.,
RA   JUJI T., KOBAYASHI N., KOHSAKA T.;
RL   MOL. IMMUNOL. 30:62-62(1993).
CC   -!- FUNCTION: C3 PLAYS A CENTRAL ROLE IN THE ACTIVATION OF THE
CC       COMPLEMENT SYSTEM. ITS PROCESSING BY C3 CONVERTASE IS THE CENTRAL
CC       REACTION IN BOTH CLASSICAL AND ALTERNATIVE COMPLEMENT PATHWAYS.
CC       AFTER ACTIVATION C3B CAN BIND COVALENTLY, VIA ITS REACTIVE
```

```
CC       THIOLESTER, TO CELL SURFACE CARBOHYDRATES OR IMMUNE AGGREGATES.
CC   -!- SUBUNIT: C3 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 ARG
CC       RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE
CC       BOND. C3 CONVERTASE ACTIVATES C3 BY CLEAVING THE ALPHA CHAIN,
CC       RELEASING C3A ANAPHYLATOXIN & GENERATING C3B (BETA CHAIN + ALPHA'
CC       CHAIN).
CC   -!- C3B IS RAPIDLY SPLIT IN TWO POSITIONS BY FACTOR I AND A COFACTOR
CC       TO FORM IC3B (INACTIVATED C3B) AND C3F WHICH IS RELEASED.
CC   -!- IC3B IS THE SLOWLY CLEAVED (POSSIBLY BY FACTOR I) TO FORM C3C AND
CC       C3DG. OTHER PROTEASES PRODUCE OTHER FRAGMENTS SUCH AS C3D OR C3G.
CC   -!- SIMILARITY: TO C4, C5 AND ALPHA-2-MACROGLOBULIN.
CC   -!- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C3,
CC       C3A ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT
CC       INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
CC       PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
CC       BASOPHILIC LEUKOCYTES.
CC   -!- POLYMORPHISM: THERE ARE TWO ALLELES: C3S (C3 SLOW), THE MOST
CC       COMMON ALLELE IN ALL RACES AND C3F (C3 FAST), RELATIVELY FREQUENT
CC       IN CAUCASOIDS, LESS COMMON IN BLACK AMERICAN, EXTREMELY RARE IN
CC       ORIENTALS.
CC   -!- DISEASE: C3 DEFICIENCY CAUSES A SUSCEPTIBILITY TO PYOGENIC
CC       INFECTION.
CC   -!- SIMILARITY: CONTAINS ONE ANAPHYLATOXIN-LIKE DOMAIN.
DR   EMBL; K02765; G179665; -.
DR   PIR; A01257; C3HU.
DR   PIR; A27603; A27603.
DR   HSSP; P01032; 1C5A.
DR   SWISS-2DPAGE; P01024; HUMAN.
DR   MIM; 120700; -.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN.
KW   COMPLEMENT PATHWAY; COMPLEMENT ALTERNATE PATHWAY; PLASMA;
KW   INFLAMMATORY RESPONSE; GLYCOPROTEIN; SIGNAL; POLYMORPHISM;
KW   DISEASE MUTATION.
FT   SIGNAL        1     22
FT   CHAIN        23   1663       COMPLEMENT C3.
FT   CHAIN        23    667       BETA CHAIN.
FT   CHAIN       672   1663       ALPHA CHAIN.
FT   PEPTIDE     672    748       C3A ANAPHYLATOXIN.
FT   CHAIN       749   1663       C3B ALPHA' CHAIN.
FT   PEPTIDE     749    954       C3C FRAGMENT.
FT   PEPTIDE     955   1303       C3DG FRAGMENT.
FT   PEPTIDE     955   1001       C3G FRAGMENT.
FT   PEPTIDE    1002   1303       C3D FRAGMENT.
FT   PEPTIDE    1304   1320       C3F FRAGMENT.
FT   SITE        748    749       CLEAVAGE (BY C3 CONVERTASE).
FT   SITE        954    955       CLEAVAGE (BY FACTOR I) (POTENTIAL).
FT   SITE       1303   1304       CLEAVAGE (BY FACTOR I).
FT   SITE       1320   1321       CLEAVAGE (BY FACTOR I).
FT   DOMAIN      693    728       ANAPHYLATOXIN-LIKE.
FT   DOMAIN     1424   1456       PROPERDIN-BINDING.
FT   DISULFID    559    816       INTERCHAIN.
FT   DISULFID    627    662
FT   DISULFID    693    720
FT   DISULFID    694    727
FT   DISULFID    707    728
FT   DISULFID    873   1513
FT   DISULFID   1101   1158
FT   DISULFID   1358   1489
FT   DISULFID   1389   1458
FT   DISULFID   1506   1511
FT   DISULFID   1518   1590
FT   DISULFID   1537   1661
FT   DISULFID   1637   1646
FT   CARBOHYD     85     85
```



```
FT   CARBOHYD    939    939
FT   CARBOHYD   1617   1617       POTENTIAL.
FT   THIOLEST   1010   1013
FT   VARIANT     102    102       R -> G (IN ALLELE C3F).
FT   VARIANT     314    314       L -> P.
FT   VARIANT     549    549       D -> N (IN C3 DEFICIENCY; IMPAIRS
FT                                SECRETION).
FT   VARIANT    1320   1320       R -> Q (IN C3 DEFICIENCY; ALLOTYPE
FT                                C3'F02'; MAY INHIBIT IC3B SYNTHESIS).
FT   CONFLICT    681    681       D -> N (IN REF. 2).
FT   CONFLICT    700    700       E -> Q (IN REF. 2).
FT   CONFLICT   1026   1026       H -> S (IN REF. 4).
SQ   SEQUENCE   1663 AA;  187163 MW;  840B4548 CRC32;

  Query Match            19.6%;  Score 2361;  DB 2;  Length 1663;
  Best Local Similarity  28.4%;  Pred. No. 0.00e+00;
  Matches  489;  Conservative  458;  Mismatches  660;  Indels  115;  Gaps   91;

Db        8 sll-ll-llthlplalg-spmysiitpnilrleseetmvleahdaqgdvpvtvtvhdfpg 64
            :|| :| :|  |  : |    | | :|:|:|: : | :|::            |:::  :|
Qy        2 GLLGILCFLIFLGKTWGQEQTYVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSYPD 61

Db       65 kklvlssektvltpatnhmgnvtftipanrefksekgrnkfvtvqatfgtqvve-kvvlv 123
            ||:  ||    |::  :    :  :|| : ::: :        :| ::   : :    | : :
Qy       62 KKFSYSSGHVHLSSENKFQNSAILTIQP-KQLPGGQNPVSYVYLEVV-SKHFSKSKRMPI 119

Db      124 slqsgylfiqtdktiytpgstvlyriftvnhkllpvgrtvmvnienpegipvkqdslssq 183
            : ::|:|||:||| :|||  :|  |::::|  | |  |   :: : :|||  |   | :
Qy      120 TYDNGFLFIHTDKPVYTPDQSVKVRVYSLNDDLKPAKRETVLTFIDPEGSEV--DMVEEI 177

Db      184 nqlgvlpls-wdipelvnmgqwkirayyenspqqvfstefevkeyvlpsfeviveptekf 242
            :::|:::::   ||        | | |:| | :       :: ||||||||| | | :|| | :
Qy      178 DHIGIISFPDFKIPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEP-E-Y 235

Db      243 yyi-yne-kglevtitarflygkkv-egtafvifgi-qd--geqrislpeslkripiedg 296
            :| |:: | :|:|| ||::| | | |:  :: ||| :|   :|:  :  ::    : :|
Qy      236 NFIGYKNFKNFEITIKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLING 295

Db      297 sgevvlsrkvlldgvqnlraedlvgkslyvsatvilhsgsdmvqaersgipivtspyqih 356
            ::| :      :  :     ||| | ||:: ||| :|:   :|| :|| | ||| ::
Qy      296 IAQVTFDSETAVKELSYYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLN 355

Db      357 ftktpkyfkpgmpfdlmvfvtnp-d---gs-pay-rvp-vav-q--ge-dtvqsltq-gd 404
            :  || ::|||:|: ::| | :: |   |: |        : | |  :: |   |:|:  |
Qy      356 LVATPLFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDD 415

Db      405 gvaklsinthpsqkplsitvrtkkqelseaeqatrtmqalpystvgnsnnylhlsvlrte 464
            ||| : :|   ::    | : |:|     :|:| :||    :|::||::: |  |:  :   :
Qy      416 GVASFVLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKA 475

Db      465 lrpgetlnvnfllrmdraheakiryytylimnkgrllkagrqvrepgqdlvvlplsittd 524
            |  || ||: ::  :  :    ||   |  |||::|||:::  | : : :      : :::| :
Qy      476 LLVGEHLNI-IVTPKS-PYIDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQN 533

Db      525 fipsfrlvayytligasgqrevvadsvwvdvkdscvgslvvksgqsedrqpvpgqqmtlk 584
            :|| ||: || : |       |:|:||||::: : |    | |   :    |  : ||| ::|:
Qy      534 MVPSSRLLVYYIVTGEQTA-ELVSDSVWLNIEEKCGNQLQVHLSPDADAYS-PGQTVSLN 591

Db      585 iegdhgarvvlvavdkgvfvlnkknkltqskiwdvvekadigctpgsgkdyagvfsdagl 644
            :     ::| | ||| :|: : :   |     :: : :||:|:|| :|:| : | ||  |||
Qy      592 MATGMDSWVALAAVDSAVYGVQRGAKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLAGL 651

Db      645 tftsssgqqtaqraelqcpqpaarrrrsvqltekrmdkvg-kypkel-rkccedgmrenp 702
            || :::   : :|   :   | :    | ||:   | :|::: :: ||    : :||| ||    |
```

```
Qy    652 TFLTNANADDSQENDEPCKE-ILRPRRT--L-QKKIEEIAAKYKHSVVKKCCYDGACVNN 707

Db    703 mrfscqrrtrfislgeackkvfldccnyitelrrqharashlqlarsnldediiaeeniv 762
            :|::|: |||| | | | :|| ::|| :: : |:| :: ::: :
Qy    708 DE-TCEQRAARISLGPRCIKAFTECCVVASQLRANISH-KDMQLGRLHMK-TLLPVSKPE 764

Db    763 srsefpeswlwnvedlkeppkngistklmniflkdsittweilavsmsdkkgicvadpfe 822
            || |||||||:| | | :: | : | ||:||||| ::::|: ||||||
Qy    765 IRSYFPESWLWEVH-L-VPRRK----QL-QFALPDSLTTWEIQGIGISNT-GICVADTVK 816

Db    823 vtvmqdffidlrlpysvvrneqveiravlynyrqnqelkvrvellhnpafcslattkrrh 882
            | | |::: :|||||| ||::::: :|||| : : | : ::|: : |
Qy    817 AKVFKDVFLEMNIPYSVVRGEQIQLKGTVYNYRTS-GMQFCVKMSAVEGICTSESPVIDH 875

Db    883 qqt-vt--ippk--ssls--vpyvivplktglqevevkaavyhhfisdgvrkslkvvpeg 935
            | | : : | :| | | : ::|| ||:::: | : : |:|:||||
Qy    876 QGTKSSKCVRQKVEGSSSHLVTFTVLPLEIGLHNINFSLETW--FGKEILVKTLRVVPEG 933

Db    936 irmnktvavrtldperl-gregvqkedip-padlsdqvpdtesetril-lqgtpvaqmte 992
            :: : :| |||| : | : :|| :| | | || || ||| : | |:::
Qy    934 VKRESYSGV-TLDPRGIYGTISRRKE-FPYRIPL-DLVPKTEI-KRILSVKGLLVGEILS 989

Db    993 davdaerlkhlivtpsgcgeqnmigmtptviavhyldeteqwekf-g--lekrqgaleli 1049
            : | :: | | | :| ::::: | |||: ::|: | : | :| :
Qy    990 AVLSQEGINILTHLPKGSAEAELMSVVPVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKL 1049

Db   1050 kkgytqqlafrqpssafaafvkrapstwltayvvkvfslavnliaidsqvlcgavkwlil 1109
            | | :::| : ::: ::||||||: ::|:: : : : :| :: ||:
Qy   1050 KEGMLSIMSYRNADYSYSVWKGGSASTWLTAFALRVLGQVNKYVEQNQNSICNSLLWLVE 1109

Db   1110 ekqkpdgvfqedapvihqemigglr-nnnekdmaltafvlislqeakdiceeqvnslpgs 1168
            : | :| | |:: : | | : |: : |||| :|::: | ||| |: : :
Qy   1110 NYQLDNGSFKENSQYQPIKLQGTLPVEARENSLYLTAFTVIGIRKAFDICP-LVK-IDTA 1167

Db   1169 itkagdfleanymnlqrsytvaiagyala--qmgr--lkg--pllnk-flttakdk-nr- 1219
            : || :|| | : | ::|:|::|||: :: : ::: : |:: | :: |
Qy   1168 LIKADNFLLENTLPAQSTFTLAISAYALSLGDKTHPQFRSIVSALKREALVKGNPPIYRF 1227

Db   1220 w-ed-pgkq--l-yn-----veatsyallallqlkdfdfvppvvrwlneqryyggygst 1269
            | :: |: : ||:|:||||: | |||:::| ||::||:|:: ||||: ||
Qy   1228 WKDNLQHKDSSVPNTGTARMVETTAYALLTSLNLKDINYVNPVIKWLSEEQRYGGGFYST 1287

Db   1270 qatfmvfqalaqyqkdapdhqelnldvslqlpsrsskithrihwesasllrseetkeneg 1329
            | |: :::|::| : |::|: : ::: :: : | |: | |:
Qy   1288 QDTINAIEGLTEYSL-LVKQLRLSMDIDVSYKHKGALHNYKMT-DKNFLGRPVEVLLNDD 1345

Db   1330 ftvtae-gkqgtlsvvtmyhakakdqltcnkfdlkvtikpapetekrpqdakntmilei 1388
            : |:: | | :|: | |: | : : |: | ||: : | : : :
Qy   1346 LIVSTGFGSGLATVHVTTVVHKTSTSEEVCS-FYLKIDTQDIEASHYRGYGNSDYKRIVA 1404

Db   1389 ctryr-g-dqdat-ms--ildismmtgfapdtddlkqlangvdryiskyeldkafsdrnt 1443
            |: |: : :: :: | ::|||: ||:::: :||| | :|||: :: |:: | : :
Qy   1405 CASYKPSREESSSGSSHAVMDISLPTGISANEEDLKALVEGVDQLFTDYQI-KD-G--H- 1459

Db   1444 liiyldkvshseddclafkvhqyfnveliqpgavkvyayynleesctrfyhpekedgkln 1503
            :|: |: :: |: |: |:: : |:| :: |:: || | : || || : : |:
Qy   1460 VILQLNSIPSSDFLCVRFRIFELFEVGFLSPATFTVYEYHRPDKQCTMFY-S-TSNIKIQ 1517

Db   1504 klcrdelcrcaeencf-iqksdd-kvtleerldkacepgvdyvyktrlvkvqlsndfdey 1561
            |:| |:| | :| :| | :: | | : || | : | ||| : : : | | |
Qy   1518 KVCEGAACKCVEADCGQMQEELDLTISAETRKQTACKPEIAYAYKVSITSITVENVFVKY 1577

Db   1562 imaieqtiksgsdevqvgqqrtfispikcrealkleekkhylmwglss-df-wgekpnls 1619
            :: : |:| : : ||| : | :| | ::||: | : ::
```



```
Qy   1578 KATLLDIYKTGEAVAEKDSEITFIKKVTCTNA-ELVKGRQYLIMGKEALQIKYNFSFRYI 1636

Db   1620 yiigkdtwvehwpeedecqdeenqkqcqdlgaftesmvvfgc 1661
            | : ||:| || : | : :| |:| : : ||
Qy   1637 YPLDSLTWIEYWPRDTTCSSCQAF-L-ANLDEFAEDIFLNGC 1676


RESULT    5
ID   CO3_LAMJA      STANDARD;      PRT;  1673 AA.
AC   Q00685;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   COMPLEMENT C3 PRECURSOR (CONTAINS: C3A ANAPHYLATOXIN) (FRAGMENT).
GN   C3.
OS   LAMPETRA JAPONICA (JAPANESE LAMPREY).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; PISCES;
OC   AGNATHA (CYCLOSTOMATA).
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 92251197.
RA   NONAKA M., TAKAHASHI M.;
RL   J. IMMUNOL. 148:3290-3295(1992).
CC   -!- FUNCTION: C3 PLAYS A CENTRAL ROLE IN THE ACTIVATION OF THE
CC       COMPLEMENT SYSTEM. AFTER ACTIVATION (C3B), IT CAN BIND COVALENTLY,
CC       VIA ITS REACTIVE THIOLESTER, TO CELL SURFACE CARBOHYDRATES OR
CC       IMMUNE AGGREGATES. CYCLOSTOMATES C3 APPEARS TO REPRESENT THE
CC       COMMON ANCESTOR OF MAMMALIAN C3 AND C4, SHOWING SIMILARITIES TO
CC       BOTH PROTEINS.
CC   -!- SIMILARITY: TO C4, C5 AND ALPHA-2-MACROGLOBULIN.
CC   -!- SIMILARITY: CONTAINS ONE ANAPHYLATOXIN-LIKE DOMAIN.
DR   EMBL; D10087; G222888; -.
KW   COMPLEMENT PATHWAY; PLASMA; INFLAMMATORY RESPONSE; GLYCOPROTEIN;
KW   SIGNAL.
FT   NON_TER       1      1
FT   SIGNAL       <1     13       POTENTIAL.
FT   CHAIN        14   1673       COMPLEMENT C3.
FT   CHAIN        14    653       BETA CHAIN (BY SIMILARITY).
FT   CHAIN       657   1375       ALPHA CHAIN (BY SIMILARITY).
FT   CHAIN      1379   1673       GAMMA CHAIN (BY SIMILARITY).
FT   PEPTIDE     657    732       C3A ANAPHYLATOXIN (BY SIMILARITY).
FT   DOMAIN      678    713       ANAPHYLATOXIN-LIKE.
FT   DISULFID    678    705       BY SIMILARITY.
FT   DISULFID    679    712       BY SIMILARITY.
FT   DISULFID    692    713       BY SIMILARITY.
FT   THIOLEST    986    990       BY SIMILARITY.
SQ   SEQUENCE   1673 AA;  187767 MW;  D857446F CRC32;

  Query Match            19.2%;  Score 2309;  DB 2;  Length 1673;
  Best Local Similarity  29.5%;  Pred. No. 0.00e+00;
  Matches  512;  Conservative  458;  Mismatches  636;  Indels  130;  Gaps  105;

Db      1 vl-llmsvgtsvtqdpmvllsvpsviligsdvnvlvdhaastedvsvvvraeefltkk-q 58
          :| :|: :| : |: ::| | :: :|: |::: : || : : ||
Qy      6 ILCFLIFLGKTWGQEQTYVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSYPDKKFS 65

Db     59 latqtitltqldpaiatlklgfdienpdktnsastkhhvrlv-akvesksfnkeitahal 117
          :: : |: : : : : |: | | ::: : | | || |:| :
Qy     66 YSSGHVHLSS-ENKFQNSAI-LTIQ-P-KQLPGG-QNPVSYVYLEVVSKHFSKS-K-RMP 118

Db    118 lsyrsghvvvqtdkpiytpdekvkyrmfpmnredvhripvrqsmtvdivnadgviverqi 177
          ::| :| : ::||||:||||: || |:::| :|: : | :: : :::::| |: :
Qy    119 ITYDNGFLFIHTDKPVYTPDQSVKVRVYSLN-DDL-K-PAKRETVLTFIDPEGSEVDM-V 174
```

```
Db    178 ktikatdegivdgtsftipaiskhgtwkifarmsgapninssaefdvreyilptfevkin 237
           |      ||:      | ||: :: | | | |:       :  ::| |:|:||:|| | | |:
Qy    175 EEI-D-HIGIISFPDFKIPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIE 232

Db    238 pkqrvfhind-eefvvditanyfnqelvsgta-yvryfl-en-gdvpk-lvdss--sttl 290
           |    :   ::  :| :  | | ||    :|: :   |: : : |:   |   | ::::: :| |
Qy    233 PEYNFIGYKNFKNFEITIKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTML 292

Db    291 vageglsilkkekllk-l-fpnakdllafsltikttvlssqaaeteeaelvgikivesry 348
          : | :    :   |  :| | : :  ||      |  |  ||: | ::  :||||: ||| | | |
Qy    293 INGIAQVTFDSETAVKELSYYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPY 352

Db    349 qitatktsryfkpelpyfiqvevrnadgspskevdvv--ak-vqvgsat--inpqk-m-r 401
           :      |: ::|| :|| | |:|:::          | |:   |   ::|    |   ::| | : |
Qy    353 KLNLVATPLFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTR 412

Db    402 tdsngltsfvt-ppnvnqltvtvrtderh-psneqgelvytaqkyas-a-symhid-v- 455
           | :|::|| :    |: |  |     |:||        |::|:   | |  |:| : ||: ||
Qy    413 VD-DGVASFVLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTD 471

Db    456 t-rimrlgetlnvfltakttqlnavthftymvltrgvivktnrktkesgggpsnvripit 514
           : : :|| ||::|:|:  :: :||:   |::|::|:       :  | | ::  :: ||:|
Qy    472 NHKALLVGEHLNIIVTPKSPYIDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVT 531

Db    515 pdmaprfrflayyilpg---geivadsvtvevtelcksqvslslkgrp-tlepkamltld 570
           :| |  |:| |||: |     :|:|:||| ::: | | :|: : |       : :  |     ::|:
Qy    532 QNMVPSSRLLVYYIVTGEQTAELVSDSVWLNIEEKCGNQLQVHLSPDADAYSPGQTVSLN 591

Db    571 ligepdarvgllavdqavyavnrkhrltqdrvwkametfdtgctaeggagrpgvfsdagl 630
          :     |::|:| ||| |||:| |   :    :||    :|   | || | ||    : || |||
Qy    592 MATGMDSWVALAAVDSAVYGVQRGAKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLAGL 651

Db    631 alitskglnttdrseigcpkvpsrkprqlsmlqirreaekytqef-rkccvdglkmsptg 689
          :::|:    : :: ::   |   :   |  | |  :   |     | || :   :|||  || ::
Qy    652 TFLTNANADDSQENDEPCKEIL-RPRRTLQK-KIEEIAAKYKHSVVKKCCYDGACVN-ND 708

Db    690 qgceerlkrvt-gpkecvdaflqcckkaeeyrkseslgaktvlrrndfmeldlmnedevn 748
          : ||:|   |:: ||: |: || :||   | : | :  |     :   | |     |    |:      :
Qy    709 ETCEQRAARISLGPR-CIKAFTECCVVASQLRANISHKDMQ-LGRLH-MK-TLLPVSKPE 764

Db    749 mmayfpqswgwnkyknsckygrhpqirlqlpdtittwnmqavsisktrgvcladplllvs 808
          : :|||:|| |:      : :   |    |:::  |||::|||::|:::||:| |:|:|| :
Qy    765 IRSYFPESWLWEVHLVP-R--RK-QLQFALPDSLTTWEIQGIGISNT-GICVADTVKAKV 819

Db    809 tkdffiklhlpysvkrgeqteirvilynymeesltiltemdivesicst-sk--s--gak 863
            || |: :::|||| ||||  :::  :|||       :: :    |    ||:|:|:  |      |:|
Qy    820 FKDVFLEMNIPYSVVRGEQIQLKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTK 879

Db    864 psq--kstvkgkgamvvsfpivplkigehhisirsrvygrtfgdgvq-kilrvapegv-r 919
          :|    :    |  |  :: :|:| ::|| || |:|:: | :   : ||   :   | ||| |||| |
Qy    880 SSKCVRQKVEGSSSHLVTFTVLPLEIGLHNINF-S-LETW-FGKEILVKTLRVVPEGVKR 936

Db    920 di-r--s-esrsvh-veeretffikneispdvvpnsdvltfisvkgdelaetmvncldak 974
           :       :  ::|::          |      |      |: |:||:::  ::||||  ::| :    |
Qy    937 ESYSGVTLDPRGIYGTISRRKEF-PYRIPLDLVPKTEIKRILSVKGLLVGEILSAVLSQE 995

Db    975 sisnliqiptgcgeqnmikmapttltliyldsvqeweki-g--lhrreeaigflkqgysr 1031
          :|: | ::| | :| ::: :  |  :  ||::       |: : :    |   ::          ||:|
Qy    996 GINILTHLPKGSAEAELMSVVPVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEGMLS 1055

Db   1032 elsyrkadhsyaafikrpsstwltafvvkvyslakrviivdnqelcgpvewiiknrqnsd 1091
           :|||:|| ||:       ::||||||| ::| :     :: :  :   :| :: |:: | |  :
Qy   1056 IMSYRNADYSYSVWKGGSASTWLTAFALRVLGQVNKYVEQNQNSICNSLLWLVENYQLDN 1115
```



```
Db   1092 gsyredgpvihremqgg--vggteghvsmtafiligiqqaqeycgvsvpnykqsmnravq 1149
           ||::|::          :||     | : |    : :||| :|||: | : | : | :     :: :|
Qy   1116 GSFKENSQYQPIKLQGTLPVEAREN SLYLTAFTVIGIRKAFDICPL-V-KIDTALIKADN 1173

Db   1150 flaskvsdlkrmytiaitryalalqd---peseaahsswkklenrttfeskghrywkae- 1205
           || :           :|:||: |||:| |      |:  :   |: |: |          ::     |:|| :
Qy   1174 FLLENTLPAQSTFTLAISAYALSLGDKTHPQFRSIVSALKR-EALVKGNPPIYRFWKDNL 1232

Db   1206 e-t-shvlrmsais-veataygllttylrkkdyesareivdwlteqrnygggfqstqdtil 1262
           :    |  |      ::      ||:|||:||| |    || :       :: ||:|:: ||||| ||||||
Qy   1233 QHKDSSVPNTGTARMVETTAYALLTSLNLKDINYVNPVIKWLSEEQRYGGGFYSTQDTIN 1292

Db   1263 alqamaqykmdssskel-idvqleit-spknnfekkmkiteetrfvqephkippggniti 1320
           |::::::| :       |:| : ::::::       |   : :   |:|:      |:    |     :     :: :
Qy   1293 AIEGLTEYSL--LVKQLRLSMDIDVSYKHKGALHN-YKMTDKN-FLGRPVEVLLNDDLIV 1348

Db   1321 kas-grgtftlsimsvfnkvapsskscstfdlkvtmteaddgespqgrlgwfdgkrrrrr 1379
           :: | |    |: :  :| :|| :       || | ||:     |: |     |:::: | |  : |:   :|
Qy   1349 STGFGSGLATVHVTTVVHKTSTSEEVCS-FYLKID-TQ-DI-EASHYR-G-Y-GNSDYKR 1401

Db   1380 digdeggveavyrmnmctrykprkedlssesgmtiievnmltgfipdkndliqlkesvdk 1439
           |          | |              |: ||| :|: || |: ::::::: ||: :: :||   | |:||
Qy   1402 -I-----V-A------CASYKPSREESSSGSSHAVMDISLPTGISANEEDLKALVEGVDQ 1448

Db   1440 yisnyeitdsvliiywdkvpstedycfafkikqmlrsdmiqpvtasvydyyspadkctrl 1499
           :::|:|  |: :|:  : :||::    |   |:| :::      :  |  | :||:|    |      || :
Qy   1449 LFTDYQIKDGHVILQLNSIPSSDFLCVRFRIFELFEVGFLSPATFTVYEYHRPDKQCTMF 1508

Db   1500 ynlpggyvelsplcqndlcqcvevscpakkpkfdtsitvlhrqeaacvagidyay-vgiv 1558
           |:   :   :   :  :|:      | |||  |    :     :| :|:      | ::|| :  | ||| |:|
Qy   1509 YS-TSN-IKIQKVCEGAACKCVEADCGQMQEELDLTISAETRKQTACKPEIAYAYKVSIT 1566

Db   1559 dnrtevgsfvyytvniqtviksgqdqaiqpkatrl-fivtrscdgrlgmetprqyllmgr 1617
                 |   || |    :   : |:| : |:    | : : ||    :|          :   ||||:||:
Qy   1567 SITVE-NVFVKYKATLLDIYKTG-E-AVAEKDSEITFIKKVTCTNA-ELVKGRQYLIMGK 1622

Db   1618 kg-etk-drndrfqyvldasswveqwpvdekcnqpnvqtfcaikreyefsmqiqgc 1671
           : :  | : : |: | ||: :|:| || |  |:  : |:| |     |:   : : ||
Qy   1623 EALQIKYNFSFRYIYPLDSLTWIEYWPRDTTCS--SCQAFLANLDEFAEDIFLNGC 1676
```

```
RESULT    6
ID   CO3_MOUSE      STANDARD;      PRT;  1663 AA.
AC   P01027;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   COMPLEMENT C3 PRECURSOR (HSE-MSF) (CONTAINS: C3A ANAPHYLATOXIN).
GN   C3.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85038854.
RA   FEY G.H., LUNDWALL A., WETSEL R.A., TACK B.F., DE BRUIJN M.H.L.,
RA   DOMDEY H.;
RL   PHILOS. TRANS. R. SOC. LOND., B, BIOL. SCI. 306:333-344(1984).
RN   [2]
RP   SEQUENCE OF 671-1663 FROM N.A.
RX   MEDLINE; 85054819.
RA   WETSEL R.A., LUNDWALL A., DAVIDSON F., GIBSON T., TACK B.F., FEY G.H.;
RL   J. BIOL. CHEM. 259:13857-13862(1984).
```

```
RN   [3]
RP   SEQUENCE OF 671-748 FROM N.A.
RX   MEDLINE; 83117730.
RA   DOMDEY H., WIEBAUER K., KAZMAIER M., MULLER V., ODINK K., FEY G.H.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 79:7619-7623(1982).
RN   [4]
RP   SEQUENCE OF 671-748 FROM N.A.
RX   MEDLINE; 84201365.
RA   FEY G.H., WIEBAUER K., DOMDEY H.;
RL   ANN. N.Y. ACAD. SCI. 421:307-312(1983).
RN   [5]
RP   SEQUENCE OF 1-34 FROM N.A.
RX   MEDLINE; 83117622.
RA   WIEBAUER K., DOMDEY H., DIGGELMANN H., FEY G.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 79:7077-7081(1982).
RN   [6]
RP   SEQUENCE OF 25-41 AND 749-760.
RX   MEDLINE; 93373334.
RA   HAMADA J.-I., CAVANAUGH P.G., MIKI K., NICOLSON G.L.;
RL   CANCER RES. 53:4418-4423(1993).
CC   -!- FUNCTION: C3 PLAYS A CENTRAL ROLE IN THE ACTIVATION OF THE
CC       COMPLEMENT SYSTEM. ITS PROCESSING BY C3 CONVERTASE IS THE CENTRAL
CC       REACTION IN BOTH CLASSICAL AND ALTERNATIVE COMPLEMENT PATHWAYS.
CC       AFTER ACTIVATION C3B CAN BIND COVALENTLY, VIA ITS REACTIVE
CC       THIOLESTER, TO CELL SURFACE CARBOHYDRATES OR IMMUNE AGGREGATES.
CC   -!- SUBUNIT: C3 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 ARG
CC       RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE
CC       BOND. C3 CONVERTASE ACTIVATES C3 BY CLEAVING THE ALPHA CHAIN,
CC       RELEASING C3A ANAPHYLATOXIN & GENERATING C3B (BETA CHAIN + ALPHA'
CC       CHAIN).
CC   -!- C3B IS RAPIDLY SPLIT IN TWO POSITIONS BY FACTOR I AND A COFACTOR
CC       TO FORM IC3B (INACTIVATED C3B) AND C3F WHICH IS RELEASED.
CC   -!- IC3B IS THE SLOWLY CLEAVED (POSSIBLY BY FACTOR I) TO FORM C3C AND
CC       C3DG. OTHER PROTEASES PRODUCE OTHER FRAGNENTS SUCH AS C3D OR C3G.
CC   -!- SIMILARITY: TO C4, C5 AND ALPHA-2-MACROGLOBULIN.
CC   -!- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C3,
CC       C3A ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT
CC       INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
CC       PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
CC       BASOPHILIC LEUKOCYTES.
CC   -!- SIMILARITY: CONTAINS ONE ANAPHYLATOXIN-LIKE DOMAIN.
DR   EMBL; K02782; G309122; -.
DR   EMBL; J00369; G387114; -.
DR   EMBL; J00367; G387114; JOINED.
DR   EMBL; M33032; E22244; -.
DR   PIR; A05290; C3MS.
DR   HSSP; P01032; 1C5A.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN.
KW   COMPLEMENT PATHWAY; COMPLEMENT ALTERNATE PATHWAY; PLASMA;
KW   INFLAMMATORY RESPONSE; GLYCOPROTEIN; SIGNAL.
FT   SIGNAL        1     24
FT   CHAIN        25   1663       COMPLEMENT C3.
FT   CHAIN        25    666       BETA CHAIN.
FT   CHAIN       671   1663       ALPHA CHAIN.
FT   PEPTIDE     671    748       C3A ANAPHYLATOXIN.
FT   CHAIN       749   1663       C3B (ALPHA' CHAIN).
FT   PEPTIDE     749    954       C3C FRAGMENT.
FT   PEPTIDE     955   1303       C3DG FRAGMENT.
FT   PEPTIDE     955   1001       C3G FRAGMENT.
FT   PEPTIDE    1002   1303       C3D FRAGMENT.
FT   PEPTIDE    1304   1320       C3F FRAGMENT.
FT   SITE        748    749       CLEAVAGE (BY C3 CONVERTASE).
FT   SITE       1303   1304       CLEAVAGE (BY FACTOR I).
FT   SITE       1320   1321       CLEAVAGE (BY FACTOR I).
```



```
FT   DOMAIN      693    728       ANAPHYLATOXIN-LIKE.
FT   DISULFID    559    816       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    626    661       BY SIMILARITY.
FT   DISULFID    693    720       BY SIMILARITY.
FT   DISULFID    694    727       BY SIMILARITY.
FT   DISULFID    707    728       BY SIMILARITY.
FT   DISULFID    873   1513       BY SIMILARITY.
FT   DISULFID   1101   1158       BY SIMILARITY.
FT   DISULFID   1358   1489       BY SIMILARITY.
FT   DISULFID   1389   1458       BY SIMILARITY.
FT   DISULFID   1506   1511       BY SIMILARITY.
FT   DISULFID   1518   1590       BY SIMILARITY.
FT   DISULFID   1537   1661       BY SIMILARITY.
FT   DISULFID   1637   1646       BY SIMILARITY.
FT   CARBOHYD    939    939
FT   CARBOHYD   1617   1617
FT   THIOLEST   1010   1013       BY SIMILARITY.
SQ   SEQUENCE   1663 AA;  186482 MW;  336F6C83 CRC32;

  Query Match            19.1%;  Score 2296;  DB 2;  Length 1663;
  Best Local Similarity  28.2%;  Pred. No. 0.00e+00;
  Matches  479;  Conservative  447;  Mismatches 665;  Indels 106;  Gaps  84;

Db     28 ysiitpnvlrleseetivleahdaqgdipvtvtvqdflkrqvltsektvltgasghlrsv 87
          | | :|:::|: : | ||::       :  |:::  :  :    |   |  ::   :::
Qy     23 YVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSYPDKKFSYSSGHVHLSSENKFQNS 82

Db     88 sikipaskefnsdkeghkyvtvvanfgetvvekavmvsfqsgylfiqtdktiytpgstvl 147
          :|    :|:: :  :   || :          | : :::::|:|||:||| :|||  :|
Qy     83 AILTIQPKQLPGGQNPVSYVYLEVVSKHFSKSKRMPITYDNGFLFIHTDKPVYTPDQSVK 142

Db    148 yriftvdnnllpvgk-tvvilietpdgipvkrdilssnnqhgilpls-wnipelvnmgqw 205
           |:::::::| |  : ||: :|: |:|  |  |::    :: ||::::  :||     | |
Qy    143 VRVYSLNDDLKPAKRETVLTFID-PEGSEV--DMVEEIDHIGIISFPDFKIPSNPRYGMW 199

Db    206 kirafyehapkqifsaefevkeyvlpsfevrveptetfy-yiddpnglevsiiakflygk 264
           |:| |        :| ||||||||| | | :||   |  |  : : :|::| |:::| |
Qy    200 TIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEPEYNFIGY-KNFKNFEITIKARYFYNK 258

Db    265 nv-dgtafvifgv-qd--gdkkislahsltrvviedgvgdavltrkvlmegvrpsnadal 320
           | ::  :: ||: :|   | |  :  ::   :: :|:::  :      : :   : : |
Qy    259 VVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLINGIAQVTFDSETAVKELSYYSLEDL 318

Db    321 vgkslyvsvtvilhsgsdmveaersgipivtspyqihftktpkffkpampfdlmvfvtnp 380
            | ||::||||  :|:   ||| :||  | ||| :::  || |:||::|: ::| | ::
Qy    319 NNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLNLVATPLFLKPGIPYPIKVQVKDS 378

Db    381 -d---gs-pas-kvl-v-vtqg-sn---akaltq-ddgvaklsintpnsrqpltitvrtk 427
           |   |: |   :   : | |  |:   :|::|: ||||| : :| |::   | : |:|
Qy    379 LDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDDGVASFVLNLPSGVTVLEFNVKTD 438

Db    428 kdtlpesrqatktmeahpystmhnsnnylhlsvsrmelkpgdnlnvnfhlrtdpgheaki 487
             |||  ||     | :||::  |  |:   : ::: |  |::||:    :: |    ||
Qy    439 APDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKALLVGEHLNIIVTPKS-P-YIDKI 496

Db    488 ryytylvmnkgkllkagrqvrepgqdlvvlslpitpefipsfrlvayytligasgqrevv 547
            | ||:::|||::  | : : :      :::|:| : :|| ||: || : |     |:|
Qy    497 THYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQNMVPSSRLLVYYIVTGEQTA-ELV 555

Db    548 adsvwvdvkdscigtlvvkgdprdnhlapgqqttlriegnqgarvglvavdkgvfvlnkk 607
          :||||::: : |   | |   |  :  :|||  :| :     ::|:| ||| :|: : :
Qy    556 SDSVWLNIEEKCGNQLQVHLSPDADAYSPGQTVSLNMATGMDSWVALAAVDSAVYGVQRG 615

Db    608 nkltqskiwdvvekadigctpgsgknyagvfmdaglafktsqglqteqradlectkpaar 667
```

```
         |    :: : :||:|:|| :|:| | | ||  |||:| |:    :   |   |  | |    |
Qy   616 AKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLAGLTFLTNANADDSQENDEPC-KEILR 674

Db   668 rrrsvqlmerrmdkagqytdkglrkccedgmrdipmryscqrrarlitqgencikafidc 727
          ||::| : :  : |: |     ::||| ||    :     :|::||  |: |  ||||| :|
Qy   675 PRRTLQ-K-KIEEIAAKYKHSVVKKCCYDGAC-VNNDETCEQRAARISLGPRCIKAFTEC 731

Db   728 cnhitklreqhrrdhvlglarseleediipeediisrshfpqswlwtieelkepekngis 787
         |   : ||     :    : |:|  :    ::|         || ||:|||| :  |  | :
Qy   732 CVVASQLRANISHKD-MQLGRLHMK-TLLPVSKPEIRSYFPESWLWEVH-L-VP-RR--- 783

Db   788 tkvmniflkdsittweilavslsdkkgicvadpyeirvmqdffidlrlpysvvrneqvei 847
          | : : | ||:||||| ::::|:  |||||      :|   | |::: :|||||| ||:::
Qy   784 -KQLQFALPDSLTTWEIQGIGISNT-GICVADTVKAKVFKDVFLEMNIPYSVVRGEQIQL 841

Db   848 ravlfnyreqe-el--kvr-ve-l-l-hnpafcsmataknryfqtikippkssvavpyvi 900
         :: ::|||    ::  |:  || :     :| :   :|  ::  :  |:   ||  | : :
Qy   842 KGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTKSSKCVRQ-KVEGSSSHLVTFTV 900

Db   901 vplkigqqevevkaavfnhfisdgvkktlkvvpegmrinktvaihtldpeklgqggv-qk 959
          :|| || ::::          |  : : |||:||||||:: :    :: ||||  :  | : ::
Qy   901 LPLEIGLHNINFSLETW--FGKEILVKTLRVVPEGVKRESYSGV-TLDPRGI-YGTISRR 956

Db   960 vdvp-aadlsdqvpdtdsetrii-lqgspvvqmaedavdgerlkhlivtpagcgeqnmig 1017
          : |     | | || |:    ||: : |  | ::     :  | :: |    | | :| ::::
Qy   957 KEFPYRIPL-DLVPKTEI-KRILSVKGLLVGEILSAVLSQEGINILTHLPKGSAEAELMS 1014

Db  1018 mtptviavhyldqteqwekfgiek-rq-ealel-ikkgytqqlafkqpssayaafnnrpp 1074
          : |     |||:  ::|: |   :    : : |    :| |     :::: :  :|:  :   ::
Qy  1015 VVPVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEGMLSIMSYRNADYSYSVWKGGSA 1074

Db  1075 stwltayvvkvfslaanliaidshvlcgavkwlilekqkpdgvfqedgpvihqemiggfr 1134
          ||||||: ::|::    : :   : : :| :: ||: : |   :| | |::          : | :
Qy  1075 STWLTAFALRVLGQVNKYVEQNQNSICNSLLWLVENYQLDNGSFKENSQYQPIKLQGTLP 1134

Db  1135 -nakeadvsltafvlialqeardicegqvnslpgsinkageyieasymnlqrpytvaiag 1193
          :|:|  : |||| :|::: | |||    |: :   :: || :::  : :   |  :|:||::
Qy  1135 VEARENSLYLTAFTVIGIRKAFDICP-LVK-IDTALIKADNFLLENTLPAQSTFTLAISA 1192

Db  1194 yalalmnkleepy--lgkflnt-a--k-drn--r-w-e--ep-dqql-yn-----veats 1234
          |||:| :|      :     :    |:    |    |    |    |  :    |      ||:|:
Qy  1193 YALSLGDKTHPQFRSIVSALKREALVKGNPPIYRFWKDNLQHKDSSVPNTGTARMVETTA 1252

Db  1235 yallalllllkdfdsvppvvrwlneqryyggqygstqatfmvfqalaqyqtdvpdhkdlnm 1294
          ||||: | |||:: | ||::||:|:: ||||: ||| |:   :::|::|    | : : |:|
Qy  1253 YALLTSLNLKDINYVNPVIKWLSEEQRYGGGFYSTQDTINAIEGLTEYSLLVKQLR-LSM 1311

Db  1295 dvsfhlpsrssattfrllwengnllrseetkqneafsltak-gkgrgtlsvvavyhaklk 1353
          |:      :::    :::   ::   |  |:  |    |:  : :::  |  | :|:  | :| |
Qy  1312 DIDVSYKHKGALHNYKMT-DKNFLGRPVEVLLNDDLIVSTGFGSGLATVHVTTVVHKTST 1370

Db  1354 skvtckkfdlrvsirpapetakkpeeakntmfleictky-lg-dvdat-ms--ildismm 1408
          |   |   | |::   :      :   :      : : :   |:  |    :  :  ::  |   ::|||:
Qy  1371 SEEVCS-FYLKIDTQDIEASHYRGYGNSDYKRIVACASYKPSREESSSGSSHAVMDISLP 1429

Db  1409 tgfapdtkdlellasgvdryiskyemnkafsnkntliiylekishteedcltfkvhqyfn 1468
          ||:::: ||  |  |||: :: |:: |  ::    :|: |: |: ::  |: |:: : |:
Qy  1430 TGISANEEDLKALVEGVDQLFTDYQI-KD-GH---VILQLNSIPSSDFLCVRFRIFELFE 1484

Db  1469 vgliqpgsvkvysyynleesctrfyhpekddgmlsklchsemcrcaeencf-mqqsqe-k 1526
          ||:: |::  || |    :  || ||  :    : :: |:| :   |:| | :| ||: :
Qy  1485 VGFLSPATFTVYEYHRPDKQCTMFY-S-TSNIKIQKVCEGAACKCVEADCGQMQEELDLT 1542

Db  1527 inlnvrldkacepgvdyvykteltnikllddfdeytmtiqqviksgsdevqagqqrkfis 1586
```



```
          |: : | : || | : | ||  :|:| : : |   |  |: :: |:|     :    :  ||
Qy  1543 ISAETRKQTACKPEIAYAYKVSITSITVENVFVKYKATLLDIYKTGEAVAEKDSEITFIK 1602

Db  1587 hikcrnalklqkgkkylmwglss-dl-wgekpntsyiigkdtwvehwpeaeecqdqkyqk 1644
          : | || | ||: ||: | : ::            | :   ||:| ||     |      |
Qy  1603 KVTCTNA-ELVKGRQYLIMGKEALQIKYNFSFRYIYPLDSLTWIEYWPRDTTC-S-SCQA 1659

Db  1645 qceelgaftesmvvygc 1661
          :|  |:| : : ||
Qy  1660 FLANLDEFAEDIFLNGC 1676
```

```
RESULT   7
ID   CO3_CAVPO      STANDARD;      PRT;  1666 AA.
AC   P12387;
DT   01-OCT-1989 (REL. 12, CREATED)
DT   01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   COMPLEMENT C3 PRECURSOR (CONTAINS: C3A ANAPHYLATOXIN).
GN   C3.
OS   CAVIA PORCELLUS (GUINEA PIG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90307998.
RA   AUERBACH H.S., BURGER R., DODDS A., COLTEN H.R.;
RL   J. CLIN. INVEST. 86:96-106(1990).
RN   [2]
RP   SEQUENCE OF 676-753.
RX   MEDLINE; 89113342.
RA   GERARD N.P., LIVELY M.O., GERARD C.;
RL   PROTEIN SEQ. DATA ANAL. 1:473-478(1988).
RN   [3]
RP   SEQUENCE OF 993-1032.
RX   MEDLINE; 83178889.
RA   THOMAS M.L., TACK B.F.;
RL   BIOCHEMISTRY 22:942-947(1983).
CC   -!- FUNCTION: C3 PLAYS A CENTRAL ROLE IN THE ACTIVATION OF THE
CC       COMPLEMENT SYSTEM. ITS PROCESSING BY C3 CONVERTASE IS THE CENTRAL
CC       REACTION IN BOTH CLASSICAL AND ALTERNATIVE COMPLEMENT PATHWAYS.
CC       AFTER ACTIVATION C3B CAN BIND COVALENTLY, VIA ITS REACTIVE
CC       THIOLESTER, TO CELL SURFACE CARBOHYDRATES OR IMMUNE AGGREGATES.
CC   -!- SUBUNIT: C3 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 ARG
CC       RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE
CC       BOND. C3 CONVERTASE ACTIVATES C3 BY CLEAVING THE ALPHA CHAIN,
CC       RELEASING C3A ANAPHYLATOXIN & GENERATING C3B (BETA CHAIN + ALPHA'
CC       CHAIN).
CC   -!- SIMILARITY: TO C4, C5 AND ALPHA-2-MACROGLOBULIN.
CC   -!- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C3,
CC       C3A ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT
CC       INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
CC       PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
CC       BASOPHILIC LEUKOCYTES.
CC   -!- SIMILARITY: CONTAINS ONE ANAPHYLATOXIN-LIKE DOMAIN.
DR   EMBL; M34054; G305335; -.
DR   PIR; A37156; A37156.
DR   PIR; S03375; S03375.
DR   PIR; D20342; D20342.
DR   HSSP; P01032; 1C5A.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN.
KW   COMPLEMENT PATHWAY; COMPLEMENT ALTERNATE PATHWAY; PLASMA;
KW   INFLAMMATORY RESPONSE; GLYCOPROTEIN; SIGNAL.
FT   SIGNAL        1     29
```

```
FT   CHAIN         30   1666       COMPLEMENT C3.
FT   CHAIN         30    671       BETA CHAIN.
FT   CHAIN        676   1666       ALPHA CHAIN.
FT   PEPTIDE      676    753       C3A ANAPHYLATOXIN.
FT   CHAIN        754   1666       C3B (ALPHA' CHAIN).
FT   SITE         753    754       CLEAVAGE (BY C3 CONVERTASE).
FT   DISULFID     557    821       INTERCHAIN (BY SIMILARITY).
FT   DISULFID     630    666       BY SIMILARITY.
FT   DOMAIN       698    733       ANAPHYLATOXIN-LIKE.
FT   DISULFID     698    725       BY SIMILARITY.
FT   DISULFID     699    732       BY SIMILARITY.
FT   DISULFID     712    733       BY SIMILARITY.
FT   DISULFID     878   1517       BY SIMILARITY.
FT   DISULFID    1106   1163       BY SIMILARITY.
FT   DISULFID    1363   1493       BY SIMILARITY.
FT   DISULFID    1394   1462       BY SIMILARITY.
FT   DISULFID    1510   1515       BY SIMILARITY.
FT   DISULFID    1522   1593       BY SIMILARITY.
FT   DISULFID    1540   1664       BY SIMILARITY.
FT   DISULFID    1640   1649       BY SIMILARITY.
FT   THIOLEST    1015   1018
FT   CARBOHYD     944    944       POTENTIAL.
FT   CARBOHYD    1620   1620       POTENTIAL.
FT   CONFLICT     731    731       D -> N (IN REF. 2).
FT   CONFLICT    1013   1013       MISSING (IN REF. 3).
FT   CONFLICT    1018   1018       Q -> E (IN REF. 2).
FT   CONFLICT    1031   1031       MISSING (IN REF. 3).
SQ   SEQUENCE    1666 AA;  186487 MW;  D8403B3F CRC32;

  Query Match            19.1%;  Score 2302;  DB 2;  Length 1666;
  Best Local Similarity  28.5%;  Pred. No. 0.00e+00;
  Matches  486;  Conservative  450;  Mismatches  658;  Indels  111;  Gaps  83;

Db     23 dpmysiitpnilrleneetvvleahevqgdipvtvtvhdfpakknvlssektvltsatgy 82
Qy     20 EQTYVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSYPDKKFSYSSGHVHLSSENKF 79

Db     83 lgtvtikipaskefksdkgr-klvvvqaafggtqlekvvlvslqsgylfiqtdktiytpg 141
Qy     80 QNSAILTIQP-KQLPGGQNPVSYVYLEVVSKHFSKSKRMPITYDNGFLFIHTDKPVYTPD 138

Db    142 stvlyriftvdsdllpvgrtiivtietpdgipikrdtlssnnqhgilpls-wnipelvnm 200
Qy    139 QSVKVRVYSLNDDLKPAKRETVLTFIDPEGSEV--DMVEEIDHIGIISFPDFKIPSNPRY 196

Db    201 gqwkiqafyenspkqvfsaefevkeyvlpsfevlveptekfy-yiddpkglevniiarfl 259
Qy    197 GMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEPEYNFIGY-KNFKNFEITIKARYF 255

Db    260 ygknv-dgtafvifgv-qd--gdqrislaqsltrvviedgsgevvlsrqvlldgvqpsrp 315
Qy    256 YNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLINGIAQVTFDSETAVKELSYYSL 315

Db    316 ealvgkslyvsvtvilhsgsdmveaersgipivtspyqihftktpkyfkpampfeimvlv 375
Qy    316 EDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLNLVATPLFLKPGIPYPIKVQV 375

Db    376 tnp-d---gs-pap-hvp-v-vtqg-snvq---sltqad-gvarlsintpntrqplsvtv 422
Qy    376 KDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDDGVASFVLNLPSGVTVLEFNV 435

Db    423 qtkkggipdarqaintmqalpyttmynsnnylhlsmprtelkpgetinvnfhlrsdpnqe 482
Qy    436 KTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKALLVGEHLNIIVTPKS-PYID 494
```



```
Db    483 akiryytylimnkgkllkvgrqprepgqalvvlpmpitkelipsfrlvayytligasaqr 542
Qy    495 -KITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQNMVPSSRLLVYYIVTGEQTA- 552

Db    543 evvadsvwadvrdscvgtlvvkggsgkdgqdkrqqhlprqqmtlriegnqgarvglvavd 602
Qy    553 ELVSDSVWLNIEEKCGNQLQVHLSP--DA-D---AYSPGQTVSLNMATGMDSWVALAAVD 606

Db    603 kgvfvlnkkhkltqskiwdvvekadigctpgsgkdyagvftdaglsfksskaglqtaqre 662
Qy    607 SAVYGVQRGAKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLAGLTFLTN-ANADDSQEN 665

Db    663 gldcpkpaarrrrsvqlmerrmdkagkykskelrrccedgmrenpmqfscqrraryvslg 722
Qy    666 DEPC-KEILRPRRTLQ-K-KIEEIAAKYKHSVVKKCCYDGACVNNDE-TCEQRAARISLG 721

Db    723 eacvkafldcctymaqlrqqhrreqnlglarsdmdediipeediisrsqfpeswlwtiee 782
Qy    722 PRCIKAFTECCVVASQLRA-NISHKDMQLGRLHMK-TLLPVSKPEIRSYFPESWLWEVH- 778

Db    783 lkeperngistktmniflkdsittweilavslsdkkgicvadpfevtvmqdffidlrlpy 842
Qy    779 L-VP-RR----KQLQFALPDSLTTWEIQGIGISNT-GICVADTVKAKVFKDVFLEMNIPY 831

Db    843 svvrneqveiravlynyrea--q-slkvr-vellhnpafcslatakkrhtqtv--tigpk 896
Qy    832 SVVRGEQIQLKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTKSSKCVRQKVEGS 891

Db    897 ssvavpyvlvplkiglqevevkaavynyfisdgvkktlkvvpegmrvnktvairtlnpeq 956
Qy    892 SSHLVTFTVLPLEIGLHNINF--SLETWFGKEILVKTLRVVPEGVKRESYSGV-TLDPRG 948

Db    957 lgqggv-qreeip-aadlsdqvpdtdsetkil-lqgtpvaqmaedavdaerlkhliitps 1013
Qy    949 I-YGTISRRKEFPYRIPL-DLVPKTEI-KRILSVKGLLVGEILSAVLSQEGINILTHLPK 1005

Db   1014 gcgeqnmigmtptviavhyldqteqwekf-g--lekrqealnlinrgytqqlafkqpnwa 1070
Qy   1006 GSAEAELMSVVPVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEGMLSIMSYRNADYS 1065

Db   1071 yaafknrasstwltayvvkvfslaanligidsevlcgavkwlilekqkpdgvfqedgpvi 1130
Qy   1066 YSVWKGGSASTWLTAFALRVLGQVNKYVEQNQNSICNSLLWLVENYQLDNGSFKENSQYQ 1125

Db   1131 hqemiggvrt-aqeadvsltafvlialqeakdicraqvnnleaninkagdyiesryadvr 1189
Qy   1126 PIKLQGTLPVEARENSLYLTAFTVIGIRKAFDIC-PLVK-IDTALIKADNFLLENTLPAQ 1183

Db   1190 rpytlaiagyalaller----lng--atlqk-fl-naateknr-weea-rqk---l-ys- 1234
Qy   1184 STFTLAISAYALSLGDKTHPQFRSIVSALKREALVKGNPPIYRFWKDNLQHKDSSVPNTG 1243

Db   1235 ----veatsyallallllkdfdavppvvrwlneqryygrgygstqatfmvfqalaqyqtd 1290
Qy   1244 TARMVETTAYALLTSLNLKDINYVNPVIKWLSEEQRYGGGFYSTQDTINAIEGLTEYSLL 1303

Db   1291 vpdhkdlnmevalqlpsrsspskfrlvweagsllrseatkqnegfkltak-gkgqgtlsv 1349
Qy   1304 VKQLR-LSMDIDVSYKHKGALHNYKMT-DKNFLGRPVEVLLNDDLIVSTGFGSGLATVHV 1361

Db   1350 vavyyaktkrkvvcknfdlrvtlkpapdtvkkpqeakstmilgictry-lg-dqdat-ms 1406
Qy   1362 TTVVHKTSTSEEVC-SFYLKIDTQDIEASHYRGYGNSDYKRIVACASYKPSREESSSGSS 1420
```

```
Db    1407 --ildismmtgfipdtddlkllatgvdryiskyemnkdfskntliiyldkvshseeecls 1464
             ::|||: ||: :: :||| |  |||: :: |:: || : : :|: |: :: |:  |:
Qy    1421 HAVMDISLPTGISANEEDLKALVEGVDQLFTDYQI-KD-G-H-VILQLNSIPSSDFLCVR 1476

Db    1465 fkihqffnvgliqpgsvkvysyynldetctqfyhpekedgmlnklchkdlcrcaeencf- 1523
             |:| ::|:||:: |:: || |   |  || || :   : :: |:|    |:| | :|
Qy    1477 FRIFELFEVGFLSPATFTVYEYHRPDKQCTMFY-S-TSNIKIQKVCEGAACKCVEADCGQ 1534

Db    1524 iq--lpekitlderlekacepgvdyvyktkllkmelsddfdeyimtieqviksgsdevqa 1581
             :|  |   |: : | : || | : | ||  :  : : : |  |  |: :: |:|    :
Qy    1535 MQEELDLTISAETRKQTACKPEIAYAYKVSITSITVENVFVKYKATLLDIYKTGEAVAEK 1594

Db    1582 gkerrfishikcrdalhlkegkhylmwglss-dl-wgerpnmsyiigkdtwveawpeaee 1639
               | ||  : | :|  |  |::||: |  : ::          | :   ||:| ||
Qy    1595 DSEITFIKKVTCTNA-ELVKGRQYLIMGKEALQIKYNFSFRYIYPLDSLTWIEYWPRDTT 1653

Db    1640 cqdeenqqqcqdlgtftenmvvfgc 1664
              |    |    :|  |:|:: : ||
Qy    1654 C-SS-CQAFLANLDEFAEDIFLNGC 1676


RESULT    8
ID   CO3_RAT        STANDARD;      PRT;   1663 AA.
AC   P01026;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   COMPLEMENT C3 PRECURSOR (CONTAINS: C3A ANAPHYLATOXIN).
GN   C3.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WISTAR; TISSUE=LIVER;
RX   MEDLINE; 90245672.
RA   MISUMI Y., SOHDA M., IKEHARA Y.;
RL   NUCLEIC ACIDS RES. 18:2178-2178(1990).
RN   [2]
RP   SEQUENCE OF 671-748.
RX   MEDLINE; 79062262.
RA   JACOBS J.W., RUBIN J.S., HUGLI T.E., BOGARDT R.A., MARIZ I.K.,
RA   DANIELS J.S., DAUGHADAY W.H., BRADSHAW R.A.;
RL   BIOCHEMISTRY 17:5031-5038(1978).
RN   [3]
RP   SEQUENCE OF 1316-1595 FROM N.A.
RX   MEDLINE; 89380332.
RA   SUNDSTROM S.A., KOMM B.S., PONCE-DE-LEON H., YI Z., TEUSCHER C.,
RA   LYTTLE C.R.;
RL   J. BIOL. CHEM. 264:16941-16947(1989).
CC   -!- FUNCTION: C3 PLAYS A CENTRAL ROLE IN THE ACTIVATION OF THE
CC       COMPLEMENT SYSTEM. ITS PROCESSING BY C3 CONVERTASE IS THE CENTRAL
CC       REACTION IN BOTH CLASSICAL AND ALTERNATIVE COMPLEMENT PATHWAYS.
CC       AFTER ACTIVATION C3B CAN BIND COVALENTLY, VIA ITS REACTIVE
CC       THIOLESTER, TO CELL SURFACE CARBOHYDRATES OR IMMUNE AGGREGATES.
CC   -!- SUBUNIT: C3 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 ARG
CC       RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE
CC       BOND. C3 CONVERTASE ACTIVATES C3 BY CLEAVING THE ALPHA CHAIN,
CC       RELEASING C3A ANAPHYLATOXIN & GENERATING C3B (BETA CHAIN + ALPHA'
CC       CHAIN).
CC   -!- SIMILARITY: TO C4, C5 AND ALPHA-2-MACROGLOBULIN.
CC   -!- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C3,
CC       C3A ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT
```



```
CC       INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
CC       PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
CC       BASOPHILIC LEUKOCYTES.
CC   -!- SIMILARITY: CONTAINS ONE ANAPHYLATOXIN-LIKE DOMAIN.
DR   EMBL; X52477; G56954; -.
DR   EMBL; M29866; E31168; ALT_SEQ.
DR   PIR; A01260; A01260.
DR   PIR; S15764; S15764.
DR   HSSP; P01032; 1C5A.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN.
KW   COMPLEMENT PATHWAY; COMPLEMENT ALTERNATE PATHWAY; PLASMA;
KW   INFLAMMATORY RESPONSE; GLYCOPROTEIN; SIGNAL.
FT   SIGNAL        1     24
FT   CHAIN        25   1663       COMPLEMENT C3.
FT   CHAIN        25    666       BETA CHAIN.
FT   CHAIN       671   1663       ALPHA CHAIN.
FT   PEPTIDE     671    748       C3A ANAPHYLATOXIN.
FT   CHAIN       749   1663       C3B (ALPHA' CHAIN).
FT   SITE        748    749       CLEAVAGE (BY C3 CONVERTASE).
FT   DOMAIN      693    728       ANAPHYLATOXIN-LIKE.
FT   DISULFID    558    816       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    626    661       BY SIMILARITY.
FT   DISULFID    693    720       BY SIMILARITY.
FT   DISULFID    694    727       BY SIMILARITY.
FT   DISULFID    707    728       BY SIMILARITY.
FT   DISULFID    873   1513       BY SIMILARITY.
FT   DISULFID   1101   1158       BY SIMILARITY.
FT   DISULFID   1358   1489       BY SIMILARITY.
FT   DISULFID   1389   1458       BY SIMILARITY.
FT   DISULFID   1506   1511       BY SIMILARITY.
FT   DISULFID   1518   1590       BY SIMILARITY.
FT   DISULFID   1537   1661       BY SIMILARITY.
FT   THIOLEST   1010   1013
FT   CARBOHYD    939    939       PROBABLE.
FT   CARBOHYD   1617   1617       PROBABLE.
FT   CONFLICT    721    722       LK -> KL (IN REF. 2).
SQ   SEQUENCE   1663 AA;  186460 MW;  0428CF63 CRC32;

  Query Match            18.7%;  Score 2255;  DB 2;  Length 1663;
  Best Local Similarity  28.7%;  Pred. No. 0.00e+00;
  Matches  478;  Conservative  422;  Mismatches  665;  Indels  102;  Gaps  76;

Db     28 ysiitpnvlrleseetfileahdaqgdvpvtvtvqdfl-kkqvltsektvltgatghlnr 86
          | | :|:::|: : | ::::              |::: :  ||  :|    |::     |
Qy     23 YVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSYPDKKFSYSSGHVHLSSENKFQNS 82

Db     87 vfikipaskefnadkgh-kyvtvvanfgatvvekavlvsfqsgylfiqtdktiytpgstv 145
          :: | : |:: :       || :            | : ::::|:|||:||| :||| :|
Qy     83 AILTIQP-KQLPGGQNPVSYVYLEVVSKHFSKSKRMPITYDNGFLFIHTDKPVYTPDQSV 141

Db    146 fyriftvdnnllpvgktvvivietpdgvpikrdilsshnqygilpls-wnipelvnmgqw 204
            |:::::::| |  :  |: :  |:|  :  |::    :: ||:::: :||     | |
Qy    142 KVRVYSLNDDLKPAKRETVLTFIDPEGSEV--DMVEEIDHIGIISFPDFKIPSNPRYGMW 199

Db    205 kirafyehapkqtfsaefevkeyvlpsfevlveptekfy-yihgpkglevsitarflygk 263
           |:| |      | :| ||||||||| | | :||  :|  | :  | :|::| ||::| |
Qy    200 TIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEPEYNFIGYKNF-KNFEITIKARYFYNK 258

Db    264 nv-dgtafvifgv-qd--edkkislalsltrvliedgsgeavlsrkvlmdgvrpsspeal 319
           | ::  :: ||: :|  :| |  :  ::   :: :| :: :      :  :   | | |
Qy    259 VVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLINGIAQVTFDSETAVKELSYYSLEDL 318

Db    320 vgkslyvsvtvilhsgsdmveaersgipivtspyqihftktpkffkpampfdlmvfvtnp 379
           | ||::|||| :|:   ||| :|| | ||| :::  || |:||::|: ::| | ::
```

```
Qy    319 NNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLNLVATPLFLKPGIPYPIKVQVKDS 378

Db    380 -d---gs-par-rvp-v-vtqg-sd---aqaltq-ddgvaklsvntpnnrqpltitvstk 426
          |    |: |      : | |  ||    : ::|: ||||| : :| |:     | : | |
Qy    379 LDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDDGVASFVLNLPSGVTVLEFNVKTD 438

Db    427 kegipdarqatrtmqaqpystmhnsnnylhlsvsrvelkpgdnlnvnfhlrtdagqeaki 486
            :|:  ||    :| :||::  |  |:  : ::  |  |::||: : :    :    ||
Qy    439 APDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKALLVGEHLNI-I-VTPKSPYIDKI 496

Db    487 ryytylvmnkgkllkagrqvrepgqdlvvlslpitpefipsfrlvayytliganqqrevv 546
           | ||:::|||::   | : : :      :::|:| : :|| ||: || : |     |:|
Qy    497 THYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQNMVPSSRLLVYYIVTGEQTA-ELV 555

Db    547 adsvwvdvkdscvgtlvvkgdprdnrqpapghqttlriegnqgarvglvavdkgvfvlnk 606
          :||||::: : |     | |     | |      :||:  :| :      ::|:| ||| :|: : :
Qy    556 SDSVWLNIEEKCGNQLQVHLSP-DADAYSPGQTVSLNMATGMDSWVALAAVDSAVYGVQR 614

Db    607 knkltqskiwdvvekadiqctpgsgknyagvfmdagltfktnqqlqtdqredpecakpaa 666
           |      :: : :||:|:|| :|:| | | ||  |||| ||    :   | :|  | |
Qy    615 GAKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLAGLTFLTNANADDSQENDEPC-KEIL 673

Db    667 rrrrsvqlmerrmdkagqytdkglrkccedgmrdipmpyscqrrarlitqgesclkafmd 726
           | ||::| : :  : |: |      ::||| ||    :      :|::||  |: |  |:||| :
Qy    674 RPRRTLQ-K-KIEEIAAKYKHSVVKKCCYDGAC-VNNDETCEQRAARISLGPRCIKAFTE 730

Db    727 ccnyitklreqhrrdhvlglarsdvdediipeediisrshfpeswlwtieelkepekngi 786
          ||   : ||      : |:|  :    ::|         || |||||| :  |  | :
Qy    731 CCVVASQLRANISHKD-MQLGRLHMK-TLLPVSKPEIRSYFPESWLWEVH-L-VP-RR-- 783

Db    787 stkvmniflkdsittweilavslsdkkgicvadpyeitvmqdffidlrlpysvvrneqve 846
            | : : | ||:|||| ::::|:  |||||       |  | |::: :|||||| ||::
Qy    784 --KQLQFALPDSLTTWEIQGIGISNT-GICVADTVKAKVFKDVFLEMNIPYSVVRGEQIQ 840

Db    847 iravlfnyreqek---lkvr-vellhnpafcsmatakkryyqti-e-ippkssvavpyvi 900
          ::: ::|||   :    :|:  || :  :      :      :     : : :    ||  | : :
Qy    841 LKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTKSSKCVRQKVEGSSSHLVTFTV 900

Db    901 vplkiglqevevkaavfnhfisdgvkkilkvvpegmrvnktvavrtldpehlnqggvqre 960
          :|| |||::::          |  : : | |:||||:: :    :| ||||   :   | : |
Qy    901 LPLEIGLHNINFSLETW--FGKEILVKTLRVVPEGVKRESYSGV-TLDPRGIY-GTISRR 956

Db    961 dvnaadls-dqvpdtdsetril-lqgtpvaqmaedavdgerlkhlivtpsgcgeqnmigm 1018
             :  :: | || |:    ||| : |  |:::      :   | :: |     | | :| :::::
Qy    957 KEFPYRIPLDLVPKTEI-KRILSVKGLLVGEILSAVLSQEGINILTHLPKGSAEAELMSV 1015

Db   1019 tptviavhyldqteqwekf-g--lekrqealelikkgytqqlafkqpisayaafnnrpps 1075
           |      |||:  ::|: | :   |  :|       :| |        :::: :  :|:  :  ::|
Qy   1016 VPVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEGMLSIMSYRNADYSYSVWKGGSAS 1075

Db   1076 twltamwsrsfslaanliaidsqvlcgavkwlilekqkpdgvfqedgpvihqemiggfr- 1134
           |||||    | ::     : :  :    :| :: ||: : |    :| | |::       : | :
Qy   1076 TWLTAFALRVLGQVNKYVEQNQNSICNSLLWLVENYQLDNGSFKENSQYQPIKLQGTLPV 1135

Db   1135 ntkeadvsltafvlialqeardicegqvnslpgsinkageyleasylnlqrpytvaiagy 1194
          :::|  : |||| :|::: | |||    |: :  :   :: || ::|    : |   |  :|:|||::|
Qy   1136 EARENSLYLTAFTVIGIRKAFDICP-LVK-IDTALIKADNFLLENTLPAQSTFTLAISAY 1193

Db   1195 alalmnkleepy--ltkflnt-a--k-drn--r-weepgqq----l-yn-----veatsy 1235
          ||:| :|      :    :   |:  |  |  :    |  |  :|:         ||:|:|
Qy   1194 ALSLGDKTHPQFRSIVSALKREALVKGNPPIYRFWKDNLQHKDSSVPNTGTARMVETTAY 1253

Db   1236 allallllkdfdsvppvvrwlnderyygggygstqatfmvfqalaqyradvpdhkdlnmd 1295
          |||: | |||:: | ||::||::|: ||||: ||| |:   :::|::|     | : : |:||
```



```
Qy   1254 ALLTSLNLKDINYVNPVIKWLSEEQRYGGGFYSTQDTINAIEGLTEYSLLVKQLR-LSMD 1312

Db   1296 vslhlpsrssptvfrllwesgsllrseetkqnegfsltak-gkgqgtlsvvtvyhakvkg 1354
             : :     :::    :::   :     |  |: |    |: : :::   | |  :|: | || |      :
Qy   1313 IDVSYKHKGALHNYKMT-DKNFLGRPVEVLLNDDLIVSTGFGSGLATVHVTTVVHKTSTS 1371

Db   1355 kttckkfdlrvtikpapetakkpqdakssmildictry-lg-dvdat-ms--ildismmt 1409
                | | |::          :   :          : :   |: |   : :  ::   |   ::|||: |
Qy   1372 EEVCS-FYLKIDTQDIEASHYRGYGNSDYKRIVACASYKPSREESSSGSSHAVMDISLPT 1430

Db   1410 gfipdtndlellssgvdryiskyemdkafsnkntliiylekishseedclsfkvhqffnv 1469
              |: :: :||   |   |||: :: |::   |    ::     :|: |: |: |:   |: |:: ::|:|
Qy   1431 GISANEEDLKALVEGVDQLFTDYQI-KD-GH---VILQLNSIPSSDFLCVRFRIFELFEV 1485

Db   1470 gliqpgsvkvysyynleesctrfyhpekddgmlsklchnemcrcaeencf-mhqsqd-qv 1527
              |:: |::   || |      :   || ||  :     : :: |:|       |:| | :|   |::   |   :
Qy   1486 GFLSPATFTVYEYHRPDKQCTMFY-S-TSNIKIQKVCEGAACKCVEADCGQMQEELDLTI 1543

Db   1528 slnerldkacepgvdyvykrtkltieliddfdeyimtieqviksgsdevqaqqerrfish 1587
             | : | : || |  : | ||   :|:|  : : |   |   |: :: |:|      :      |   ||
Qy   1544 SAETRKQTACKPEIAYAYKVSITSITVENVFVKYKATLLDIYKTGEAVAEKDSEITFIKK 1603

Db   1588 vkcrnalklqkgkqylmwglss-dl-wgekpntsyiigkdtwvehwp 1632
             | | ||  | ||:|||: |  : ::                |  :    ||:| ||
Qy   1604 VTCTNA-ELVKGRQYLIMGKEALQIKYNFSFRYIYPLDSLTWIEYWP 1649
```

```
RESULT    9
ID   CO3_EPTBU      STANDARD;      PRT;  1620 AA.
AC   P98094;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   COMPLEMENT C3 (CONTAINS: C3A ANAPHYLATOXIN) (FRAGMENT).
GN   C3.
OS   EPTATRETUS BURGERI (INSHORE HAGFISH).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; PISCES;
OC   AGNATHA (CYCLOSTOMATA).
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE; 92192016.
RA   ISHIGURO H., KOBAYASHI K., SUZUKI M., TITANI K., TOMONAGA S.,
RA   KUROSAWA Y.;
RL   EMBO J. 11:829-837(1992).
CC   -!- FUNCTION: C3 PLAYS A CENTRAL ROLE IN THE ACTIVATION OF THE
CC       COMPLEMENT SYSTEM. AFTER ACTIVATION (C3B), IT CAN BIND COVALENTLY,
CC       VIA ITS REACTIVE THIOLESTER, TO CELL SURFACE CARBOHYDRATES OR
CC       IMMUNE AGGREGATES. CYCLOSTOMATES C3 APPEARS TO REPRESENT THE
CC       COMMON ANCESTOR OF MAMMALIAN C3 AND C4, SHOWING SIMILARITIES TO
CC       BOTH PROTEINS.
CC   -!- SIMILARITY: TO C4, C5 AND ALPHA-2-MACROGLOBULIN.
CC   -!- SIMILARITY: CONTAINS ONE ANAPHYLATOXIN-LIKE DOMAIN.
DR   EMBL; Z11595; G62775; -.
DR   EMBL; Z11596; -; NOT_ANNOTATED_CDS.
DR   PIR; S21045; S21045.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN.
KW   COMPLEMENT PATHWAY; PLASMA; INFLAMMATORY RESPONSE; GLYCOPROTEIN;
KW   SIGNAL.
FT   NON_TER       1      1
FT   CHAIN        <1   1620       COMPLEMENT C3.
FT   CHAIN        <1    633       BETA CHAIN (BY SIMILARITY).
FT   CHAIN       634   1336       ALPHA CHAIN (BY SIMILARITY).
FT   CHAIN      1343   1620       GAMMA CHAIN (BY SIMILARITY).
FT   PEPTIDE     634    714       C3A ANAPHYLATOXIN (BY SIMILARITY).
```

```
FT   DOMAIN      658    694       ANAPHYLATOXIN-LIKE.
FT   DOMAIN     1406   1416       PROPERDIN-BINDING.
FT   DISULFID    526    780       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    593    628       BY SIMILARITY.
FT   DISULFID    658    686       BY SIMILARITY.
FT   DISULFID    659    693       BY SIMILARITY.
FT   DISULFID    672    694       BY SIMILARITY.
FT   DISULFID    836   1474       BY SIMILARITY.
FT   DISULFID   1062   1114       BY SIMILARITY.
FT   DISULFID   1321   1451       BY SIMILARITY.
FT   DISULFID   1356   1420       BY SIMILARITY.
FT   DISULFID   1467   1472       BY SIMILARITY.
FT   DISULFID   1479   1551       BY SIMILARITY.
FT   DISULFID   1499   1616       BY SIMILARITY.
FT   DISULFID   1598   1607       BY SIMILARITY.
FT   THIOLEST    970    973       BY SIMILARITY.
SQ   SEQUENCE   1620 AA;  181747 MW;  C3B9D702 CRC32;

  Query Match            17.5%;  Score 2106;  DB 2;  Length 1620;
  Best Local Similarity  29.6%;  Pred. No. 0.00e+00;
  Matches  466;  Conservative  416;  Mismatches  588;  Indels  104;  Gaps   88;

Db      93 ipltsrrglvfaqtdqpiytpnndvnirlfpvtrqlnpilsslvvdimnpdgvvvdriek 152
Qy     117 MPITYDNGFLFIHTDKPVYTPDQSVKVRVYSLNDDLKPAKRETVLTFIDPEGSEVDMVEE 176

Db     153 nafevekvmelrpfhvpaitslgdwkivswmkdkpqfnyt-sg-fkveeyvlptfdvsit 210
Qy     177 ID-HIG-IISFPDFKIPSNPRYGMWTIKA--KYKEDFSTTGTAYFEVKEYVLPHFSVSIE 232

Db     211 seqpylhvy-d-kaftihikamhiygkpvm-grayvrygvkhqsk---rtllstssalar 264
Qy     233 PEYNFI-GYKNFKNFEITIKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTM 291

Db     265 feqqeamhtl-rqkhile-qypdpklllgqslyveasvissdaqeiensilddipivasp 322
Qy     292 LINGIAQVTFDSETAVKELSYYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSP 351

Db     323 ysikskwtvpffkpgvpyiykvlvlnpdgspasgvp-i-kv-sfsfd-ssgnwitqkrkt 378
Qy     352 YKLNLVATPLFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVT 411

Db     379 -mdngiamqtintarnskklnikvqtederleqsqqaeasftiasyssps gsfihln-a- 435
Qy     412 RVDDGVASFVLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTD 471

Db     436 hrevkspgehivfdvfiksaakdhvlhfnylmisngkihnflqegrkgdtt--svslllt 493
Qy     472 NHKALLVGEHLNIIVTPKSPYIDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVT 531

Db     494 pelvpqfrlvaffilp---sgelvadsiiidvkdschaklsldvaggkrlfsprdnvnfd 550
Qy     532 QNMVPSSRLLVYYIVTGEQTAELVSDSVWLNIEEKCGNQLQVHLSPDADAYSPGQTVSLN 591

Db     551 lsgesdswvavgvvdkaayvldkknkltankvykameasdlgcsvgsgktgplvfrdagl 610
Qy     592 MATGMDSWVALAAVDSAVYGVQRGAKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLAGL 651

Db     611 aimakeisgmddvkdpgcpnghtrrkrelvleiaiekasty-paelrkccrdaaiesplr 669
Qy     652 TFLTNANADDSQENDEPC-KEILRPRRTLQKKIE-EIAAKYKHSVVKKCCYDGACVNN-D 708

Db     670 lsceertkhihdegegcqetflecckhveeelliameeededlgrsqgedfmiqesqvvi 729
```



```
Qy     709 ETCEQRAARI-SLGPRCIKAFTECCV-VASQLRANISHKDMQLGRLHMKT-LLPVSKPEI 765

Db     730 rshfpesfmweiiklsrsaengksritkkmpdsittwdiqavevsqskglcvgpsleltv 789
Qy     766 RSYFPESWLWEVHLVPRR--K-QLQFA--LPDSLTTWEIQGIGISNT-GICVADTVKAKV 819

Db     790 fkqfflkvhtpyalkqyeqvelrvviynymnqdvkgeiqvkcgdgict-dae--qneplk 846
Qy     820 FKDVFLEMNIPYSVVRGEQIQLKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTK 879

Db     847 srfav-ek-nsats-f-sfmvvplsssdssvs-vlarvfgsdvhdavekdlrvmpegnye 901
Qy     880 SSKCVRQKVEGSSSHLVTFTVLPLEIGLHNINFSLETWFGKEI--LV-KTLRVVPEGVKR 936

Db     902 emsrswsvqpr-rhgg-qqvivvdnetpqnvvpgtemsaflsaqgnlvaetiqntlkgsk 959
Qy     937 ESYSGVTLDPRGIYGTISRRKEFPYRIPLDLVPKTEIKRILSVKGLLVGEILSAVLSQEG 996

Db     960 isnllrlprgcgeqnmmytsitvm-varylnrsdqwnkm-gdpqlkkrsfdf-itsgfas 1016
Qy     997 INILTHLPKGSAEAELM-SVVPVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEGMLS 1055

Db    1017 qltyrkpdysyaawlhrasstwltafvakvfsqarqlvfipvseicgsvrwlmrk-qdkd 1075
Qy    1056 IMSYRNADYSYSVWKGGSASTWLTAFALRVLGQVNKYVEQNQNSICNSLLWLVENYQLDN 1115

Db    1076 gsfle-sk-pvvhln-mmg-qvtgkvv-ltsfvfiallearesc-inevegftvvvekah 1129
Qy    1116 GSFKENSQYQPIKLQGTLPVEARENSLYLTAFTVIGIRKAFDICPLVKID--TALI-KAD 1172

Db    1130 gyltsqamngledfplaitayalslw-kv-sdgaakvt-m-h-tl-k-tsgl-q--teel 1179
Qy    1173 NFLLENTLPAQSTFTLAISAYALSLGDKTHPQFRSIVSALKREALVKGNPPIYRFWKDNL 1232

Db    1180 ihw-gs--nkgkaaavestaygllaaiqheegeiaekatnwlsqsatfggyfqstqdtvm 1236
Qy    1233 QHKDSSVPNTGTARMVETTAYALLTSLNLKDINYVNPVIKWLSEEQRYGGGFYSTQDTIN 1292

Db    1237 alqaltgfescqsrmkkmdlsfkiraeengvfdkefqitndnafvqkpfkvpvhgqltvt 1296
Qy    1293 AIEGLTEY-SLLVKQLRLSMDIDVSYKHKGALHN-YKMTDKN-FLGRPVEVLLNDDLIVS 1349

Db    1297 asgtgqgiltf-vkkyrekvvikkd-ckgfsleittnldnqvkqrrrqsinpefnvyrfi 1354
Qy    1350 T-GFGSGLATVHVTTVVHKTSTSEEVC-SFYLKIDTQ-DIEASHYRGYG-NSDYK--RIV 1403

Db    1355 gcfry-l-rnqep-gm--vvmdislptgfeakkkdlddmknlvdnyivqyeirpgrvfly 1409
Qy    1404 ACASYKPSREESSSGSSHAVMDISLPTGISANEEDLKALVEGVDQLFTDYQIKDGHVILQ 1463

Db    1410 ldkvnkdekncvgfrlnqvfesnlvlpvtatvfeyyepdfrcsksyhpkmevnpdaschg 1469
Qy    1464 LNSIPSSDFLCVRFRIFELFEVGFLSPATFTVYEYHRPDKQCTMFY-STSNIKIQKVCEG 1522

Db    1470 nicnclqrhcvelkgmadedrnad-rngnacrae--yvfiigvtkvtktasyininaalk 1526
Qy    1523 AACKCVEADCGQMQEELDLTISAETRKQTACKPEIAYAYKVSITSITVENVFVKYKATLL 1582

Db    1527 tvlkkgmdqainvgarrsfvipmhcgknlnvspgdiylvmgmhnahwr-nsd-rtqyvlt 1584
Qy    1583 DIYKTG-EAVAEKDSEITFIKKVTC-TNAELVKGRQYLIMGKEALQIKYNFSFRYIYPLD 1640

Db    1585 sdtwfekfplesvc 1598
```

```
Qy    1641 SLTWIEYWPRDTTC 1654


RESULT   10
ID    CO3_ONCMY      STANDARD;      PRT;  1640 AA.
AC    P98093;
DT    01-FEB-1996 (REL. 33, CREATED)
DT    01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT    01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE    COMPLEMENT C3-1 (CONTAINS: C3A ANAPHYLATOXIN) (FRAGMENT).
OS    ONCORHYNCHUS MYKISS (RAINBOW TROUT) (SALMO GAIRDNERI).
OC    EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; PISCES; GNATHOSTOMATA;
OC    OSTEICHTHYES; ACTINOPTERYGII; SALMONIFORMES.
RN    [1]
RP    SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC    TISSUE=LIVER;
RX    MEDLINE; 94065166.
RA    LAMBRIS J.D., LAO Z., PANG J., ALSENZ J.;
RL    J. IMMUNOL. 151:6123-6134(1993).
CC    -!- FUNCTION: C3 PLAYS A CENTRAL ROLE IN THE ACTIVATION OF THE
CC        COMPLEMENT SYSTEM. ITS PROCESSING BY C3 CONVERTASE IS THE CENTRAL
CC        REACTION IN BOTH CLASSICAL AND ALTERNATIVE COMPLEMENT PATHWAYS.
CC        AFTER ACTIVATION C3B CAN BIND COVALENTLY, VIA ITS REACTIVE
CC        THIOLESTER, TO CELL SURFACE CARBOHYDRATES OR IMMUNE AGGREGATES.
CC    -!- SUBUNIT: C3 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 ARG
CC        RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE
CC        BOND. C3 CONVERTASE ACTIVATES C3 BY CLEAVING THE ALPHA CHAIN,
CC        RELEASING C3A ANAPHYLATOXIN & GENERATING C3B (BETA CHAIN + ALPHA'
CC        CHAIN).
CC    -!- SIMILARITY: TO C4, C5 AND ALPHA-2-MACROGLOBULIN.
CC    -!- SIMILARITY: CONTAINS ONE ANAPHYLATOXIN-LIKE DOMAIN.
DR    EMBL; L24433; G431608; ALT_INIT.
KW    COMPLEMENT PATHWAY; COMPLEMENT ALTERNATE PATHWAY; PLASMA;
KW    INFLAMMATORY RESPONSE; GLYCOPROTEIN.
FT    NON_TER       1      1
FT    CHAIN         1   1640       COMPLEMENT C3.
FT    CHAIN         1    642       BETA CHAIN.
FT    CHAIN       647   1640       ALPHA CHAIN.
FT    PEPTIDE     647    722       C3A ANAPHYLATOXIN.
FT    CHAIN       723   1640       C3B (ALPHA' CHAIN).
FT    PEPTIDE     723    931       C3C FRAGMENT (BY SIMILARITY).
FT    PEPTIDE     932   1278       C3DG FRAGMENT (BY SIMILARITY).
FT    PEPTIDE     932   1033       C3G FRAGMENT (BY SIMILARITY).
FT    PEPTIDE    1034   1278       C3D FRAGMENT (BY SIMILARITY).
FT    PEPTIDE    1279   1295       C3F FRAGMENT (BY SIMILARITY).
FT    SITE        722    723       CLEAVAGE (BY C3 CONVERTASE).
FT    SITE        931    932       CLEAVAGE (BY FACTOR I) (BY SIMILARITY).
FT    SITE       1278   1279       CLEAVAGE (BY FACTOR I) (BY SIMILARITY).
FT    SITE       1295   1296       CLEAVAGE (BY FACTOR I).
FT    DOMAIN      668    703       ANAPHYLATOXIN-LIKE.
FT    DOMAIN     1420   1430       PROPERDIN-BINDING.
FT    DISULFID    536    797       INTERCHAIN (BY SIMILARITY).
FT    DISULFID    603    638       BY SIMILARITY.
FT    DISULFID    668    695       BY SIMILARITY.
FT    DISULFID    669    702       BY SIMILARITY.
FT    DISULFID    682    703       BY SIMILARITY.
FT    DISULFID    853   1488       BY SIMILARITY.
FT    DISULFID   1079   1135       BY SIMILARITY.
FT    DISULFID   1335   1464       BY SIMILARITY.
FT    DISULFID   1481   1486       BY SIMILARITY.
FT    DISULFID   1493   1563       BY SIMILARITY.
FT    DISULFID   1510   1638       BY SIMILARITY.
FT    DISULFID   1614   1623       BY SIMILARITY
FT    CARBOHYD    164    164
```



```
FT    THIOLEST    988    991       BY SIMILARITY.
SQ    SEQUENCE   1640 AA;  182104 MW;  F4535BB8 CRC32;

  Query Match             16.6%;  Score 2002;  DB 2;  Length 1640;
  Best Local Similarity   28.7%;  Pred. No. 0.00e+00;
  Matches  491;  Conservative  429;  Mismatches 657;  Indels 133;  Gaps  90;

Db      5 vlsapnllrvgsnenifvesqdhvggplnvkimvknhptqskelasksvvldqannfqam 64
          |:|||:::|||::||| ::       : ::  | :|: |       :|  | |   |:||
Qy     24 VISAPKIFRVGASENIVIQVYGYTEA-FDATISIKSYPDKKFSYSSGHVHLSSENKFQNS 82

Db     65 tqlviqrgplvddpkqkqyvvlqaqfpdrllekvvlvsfqsgyifiqtdktiytpastvh 124
          : | ||   |     :    || |:          | : :::::|::||:||| :|||  :|
Qy     83 AILTIQPKQLPGGQNPVSYVYLEVVSKHFSKSKRMPITYDNGFLFIHTDKPVYTPDQSVK 142

Db    125 yrvfsmtpglepltreifedqevaknkeiavsveimtpenitifreivnpdkgvksgqfk 184
           ||:|:   | |  ||          :    |: : || :  ::| |:  :  ||        ||
Qy    143 VRVYSLNDDLKPAKRETVLTFIDPEGSEVDM-VEEI--DHIGIIS-F--PD-------FK 189

Db    185 lpdivsfgtwhvvtrfqstpqktfssefevkeyvlpsfevsltpakaff-yvddndltvd 243
          :|     :| | : :::        | :: |||||||| | ||: |    |: | : ::: :
Qy    190 IPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEPEYNFIGYKNFKNFEIT 249

Db    244 itarylygkevtgtg-yvvfgvi-t-tesekksfpaslqrveikdgkgvaclkkeh-itq 299
          | |||:| | || :  |: ||:       : :|     :::|    : :| :    :  |   : :
Qy    250 IKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLINGIAQVTFDSETAVKE 309

Db    300 -tfpkihdlvkqsifvsvsvltegggemveaekrgiqivtspysilfkrtpkyfkpgmpf 358
           ::  : || :  ::::|:|:   ||    |||   ||   | |||  : :   || ::|||:|:
Qy    310 LSYYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLNLVATPLFLKPGIPY 369

Db    359 dvsvyitnp-dn---s-pai-gve-vev---tpd-h-akgvtran-gfakiplntvasat 405
           : | : :: |     : | |   : ::|    |:|    :|:||| : | | : ||   :: |
Qy    370 PIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDDGVASFVLNLPSGVT 429

Db    406 elvitvktkdpgdprqqtgggtmkalpyrtstknflhvg-vdsn-elkigdpikidlnlg 463
           | : |||  |  | :   :       :|::| : : ::| :    |::   | :|: ::| :
Qy    430 VLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKALLVGEHLNIIVTPK 489

Db    464 pttipn-hdltymflsrgqlvkvg-rfkrqgnalvtlsvpvskellpsfrivayyhv-g- 519
          :  | :       |::||:| ::  | | |      :  ::::||: :::|| |:: || | |
Qy    490 SPYIDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQNMVPSSRLLVYYIVTGE 549

Db    520 -aadlvadsvwvdikvscmgslkvtstrpkasyeprrafsltitgdpgakvglvavdkgv 578
           :|:||:||||::|     |     | |   :       :| | :: || ::       : |:| ||| :|
Qy    550 QTAELVSDSVWLNIEEKCGNQLQVHLSPDADAYSPGQTVSLNMATGMDSWVALAAVDSAV 609

Db    579 yvlnskhrltqtkiwdtiekhdtgctagggadnmgvfydaglvfetntakgtgirtdpsc 638
          | :      :      :: : :|| | || |||| :|   ||   ||| | ||:      :     | :|
Qy    610 YGVQRGAKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLAGLTFLTNANADDSQENDEPC 669

Db    639 pvssrrrravtisdvitsmasky-hglakeccvdgmrdntmgytcdrraqyisdgdvcvq 697
              | ||  |:    |   :|:|| |:: | || ||     |     ||::||  || |   |:
Qy    670 KEILRPRR--TLQKKIEEIAAKYKHSVVKKCCYDGACVNNDE-TCEQRAARISLGPRCIK 726

Db    698 aflvcctemaskkieskqdalllsrseeddddaymrsedivsrsqfpeswmwedtnlpe 757
          ||  ||   :   :  :   :  : |:|                 ::| |  || |||||:||  :|
Qy    727 AFTECCVVASQLRANISHKDMQLGRLHMKTLLPV-SKPE-I--RSYFPESWLWE-VHLV- 780

Db    758 cpaqnkhcestsvirnnflkdsittwqitaislskthgicvadpfemivlkeffidlklp 817
           | : |:       ::    | ||:|||:| :|::|:| ||||||       |:|: |:::::|
Qy    781 -P-RRKQ------LQFA-LPDSLTTWEIQGIGISNT-GICVADTVKAKVFKDVFLEMNIP 830

Db    818 ysavrneqlevkailhnysedpiivrvelmengevcssas----kkg-kyrqev--nmdp 870
```

```
                 || || ||:::|: : ||    :   | :    :|:|  |     | |    |  :::
Qy    831 YSVVRGEQIQLKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTKSSKCVRQKVEG 890

Db    871 mstrvvpyviipmklglhsievkasvknsgsndgvkrdlrvvaegvlvkketnvllnp-v 929
             |:::| : ::|: :|||:|:   |:    ::: : : ||||:|||     : | |:|
Qy    891 SSSHLVTFTVLPLEIGLHNINF--SLETWFGKEILVKTLRVVPEGVKRESYSGVTLDPRG 948

Db    930 khgg-eqtshipsgvprnqvpnsdadtlisvtageqtsvlveqaisgdslgslivqpvgc 988
               |   :   :|  :| : ||:::    ::||  |   : ::   :| :::  |   | |
Qy    949 IYGTISRRKEFPYRIPLDLVPKTEIKRILSVK-GLLVGEILSAVLSQEGINILTHLPKGS 1007

Db    989 geqnmiymtlpviat-hyldntkkwedi-g--ldkrntaikyinigyqrqlayrkedgsy 1044
             :| ::: : :||:   |||:  : |: : :  |  :     | :: |    ::||: | ||
Qy   1008 AEAELMSV-VPVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEGMLSIMSYRNADYSY 1066

Db   1045 aawvsrqsstwltayvvkvfamsstlisvqenvlctavkwlilntqqpdgifnefapvih 1104
             : | :  :||||||: ::|::  :  :   :| :| :: ||: | |  :| |:| :
Qy   1067 SVWKGGSASTWLTAFALRVLGQVNKYVEQNQNSICNSLLWLVENYQLDNGSFKENSQYQP 1126

Db   1105 aemtgn--vrgsdndasmtafvliamqeassvceqsvnslpgsmakavayl-ekrlp-h- 1159
               : |    | : :|    :||| :|::: |  :|    |: :  :: ||  :| |: || :
Qy   1127 IKLQGTLPVEARENSLYLTAFTVIGIRKAFDICPL-VK-IDTALIKADNFLLENTLPAQS 1184

Db   1160 ---lt-npyavam--tsya-l-anagklnketllkfasp------q-ldh--wpvpg-gy 1201
                |: ::||:::    : : : :   : |::|:|:|  :|        : |:|    :|| |
Qy   1185 TFTLAISAYALSLGDKTHPQFRSIVSALKREALVKGNPPIYRFWKDNLQHKDSSVPNTGT 1244

Db   1202 qytleatsyallalvkvkafeeagpivrwlnkqkkvgggygstqstimvfqavaeywshv 1261
                :|:|:||||: :::| ::    |:::||: : : |||: ||| ||  ::::::||     |
Qy   1245 ARMVETTAYALLTSLNLKDINYVNPVIKWLSEEQRYGGGFYSTQDTINAIEGLTEYSLLV 1304

Db   1262 kdlkdfdlninlevagrasvtkwsinnknqfhtrtdkvnsidkdltvkas-gngeatlsv 1320
             |:|: : ::|::     :::: :   : :|| |   |    |  :: || | :: |:| ||: |
Qy   1305 KQLR-LSMDIDVSYKHKGALHNYKMTDKN-FLGRPVEVL-LNDDLIVSTGFGSGLATVHV 1361

Db   1321 vtlyyalpeekdsdcesfdlsvtltkmdktshed-akesfmltie-vlykns--erda-t 1375
             |:    :   : | ||| |:       :: :       :: :: :      ||| |  : :
Qy   1362 TTVVHKTSTSEEV-C-SFYLKIDTQDIEASHYRGYGNSDYKRIVACASYKPSREESSSGS 1419

Db   1376 m-sildiglltgfivdtddlnqlskgreryiekfemdkvlsergslilyldkvshkledr 1434
               :::||:| ||:   : :||: |   | :: :   ::: |    : | :|| |: ::
Qy   1420 SHAVMDISLPTGISANEEDLKALVEGVDQLFTDYQI-K---D-GHVILQLNSIPSSDFLC 1474

Db   1435 isfkihrvqevgvlqpaavsvyeyyn-qkrcvkfyhpqreggtlsrlclgdvctcaeesc 1493
             : |:|  : ||| | ||: :||||   :|:| :|| :    :  ::| |  | | | |    |
Qy   1475 VRFRIFELFEVGFLSPATFTVYEYHRPDKQCTMFY-STSNI-KIQKVCEGAACKCVEADC 1532

Db   1494 s-mqkkgepdv--q-ridkacgagldyvykatvvdsk!:thtdtytvkidlvi-kpgtde 1548
             : ||    :   :   :  | : || : :  | || ::     :         |    : | | | |
Qy   1533 GQMQEELDLTISAETRKQTACKPEIAYAYKVSITSITVENVFVKYKATL-LDIYKTGEAV 1591

Db   1549 gvegknrdfmglaycrealglmqgktymimgksedlhrvedkgllqykyvlgeqtwieyw 1608
             :      :   |:     | :|  |: |: |:|||| | |: :  :  ::| | |   ||||||
Qy   1592 AEKDSEITFIKKVTCTNA-ELVKGRQYLIMGK-EALQ-IKYNFSFRYIYPLDSLTWIEYW 1648

Db   1609 psqqectsrdyrevclgidefinqittfgc 1638
             | :  |:|   :     :||| ::|   ||
Qy   1649 PRDTTCSS--CQAFLANLDEFAEDIFLNGC 1676

RESULT   11
ID    CO4_HUMAN      STANDARD;      PRT;  1741 AA.
AC    P01028;
DT    21-JUL-1986 (REL. 01, CREATED)
```



```
DT    01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT    01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE    COMPLEMENT C4 PRECURSOR (CONTAINS: C4A ANAPHYLATOXIN).
GN    C4A AND C4B.
OS    HOMO SAPIENS (HUMAN).
OC    EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC    EUTHERIA; PRIMATES.
RN    [1]
RP    SEQUENCE OF 1-22 AND 1056-1225 FROM N.A.
RX    MEDLINE; 85156269.
RA    BELT K.T., YU C.Y., CARROLL M.C., PORTER R.R.;
RL    IMMUNOGENETICS 21:173-180(1985).
RN    [2]
RP    SEQUENCE OF 20-1741 FROM N.A.
RC    TISSUE=LIVER;
RX    MEDLINE; 84156544.
RA    BELT K.T., CARROLL M.C., PORTER R.R.;
RL    CELL 36:907-914(1984).
RN    [3]
RP    SEQUENCE OF 680-756.
RX    MEDLINE; 81264286.
RA    MOON K.E., GORSKI J.P., HUGLI T.E.;
RL    J. BIOL. CHEM. 256:8685-8692(1981).
RN    [4]
RP    SEQUENCE OF 957-1044.
RX    MEDLINE; 82182029.
RA    CAMPBELL R.D., GAGNON J., PORTER R.R.;
RL    BIOCHEM. J. 199:359-370(1981).
RN    [5]
RP    SEQUENCE OF 990-1037.
RX    MEDLINE; 82150875.
RA    HARRISON R.A., THOMAS M.L., TACK B.F.;
RL    PROC. NATL. ACAD. SCI. U.S.A. 78:7388-7392(1981).
RN    [6]
RP    STRUCTURAL BASIS OF POLYMORPHISM.
RX    MEDLINE; 87080272.
RA    YU C.Y., BELT K.T., GILES C.M., CAMPBELL R.D., PORTER R.R.;
RL    EMBO J. 5:2873-2881(1986).
RN    [7]
RP    VARIANT C4A6 ALLOTYPE.
RX    MEDLINE; 92242905.
RA    ANDERSON M.J., MILNER C.M., COTTON G.H., CAMPBELL R.D.;
RL    J. IMMUNOL. 148:2795-2802(1992).
CC    -!- FUNCTION: C4 PLAYS A CENTRAL ROLE IN THE ACTIVATION OF THE
CC        CLASSICAL PATHWAY OF THE COMPLEMENT SYSTEM. IT IS PROCESSED BY
CC        ACTIVATED C1 WHICH REMOVE FROM THE ALPHA CHAIN THE C4A
CC        ANAPHYLATOXIN.
CC    -!- SUBUNIT: THIS PROTEIN IS SYNTHESIZED AS A SINGLE-CHAIN PRECURSOR
CC        AND, PRIOR TO SECRETION, IS ENZYMATICALLY CLEAVED TO FORM A TRIMER
CC        OF NONIDENTICAL CHAINS (ALPHA, BETA, AND GAMMA).
CC    -!- POLYMORPHISM: HUMAN COMPLEMENT COMPONENT C4 IS POLYMORPHIC WITH AT
CC        LEAST TWO LOCI, C4A & C4B. 13 ALLELES OF C4A & 22 ALLELES OF C4B
CC        HAVE BEEN DETECTED. THE ALLELE SHOWN HERE IS C4A4.
CC    -!- C4A ALLOTYPES REACT MORE RAPIDLY WITH THE AMINO GROUP OF PEPTIDE
CC        ANTIGENS WHILE C4B ALLOTYPES REACT MORE RAPIDLY WITH THE HYDROXYL
CC        GROUP OF CARBOHYDRATE ANTIGENS.
CC    -!- POLYMORPHISM: THE C4A ALLELES CARRY THE BLOOD GROUP RODGERS WHILE
CC        THE C4B ALLELES CARRY THE BLOOD GROUP CHIDO.
CC    -!- DISEASE: THE C4A6 ALLOTYPE IS TOTALLY DEFICIENT IN HEMOLYTIC
CC        ACTIVITY.
CC    -!- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C4,
CC        C4A ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT
CC        INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
CC        PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
```

```
CC       BASOPHILIC LEUKOCYTES.
CC   -!- C4 IS A MAJOR HISTOCOMPATIBILITY COMPLEX CLASS-III PROTEIN.
CC   -!- SIMILARITY: TO C3, C5 AND ALPHA-2-MACROGLOBULIN.
CC   -!- SIMILARITY: CONTAINS ONE ANAPHYLATOXIN-LIKE DOMAIN.
DR   EMBL; M14823; G553211; -.
DR   EMBL; K02403; G443671; ALT_SEQ.
DR   EMBL; M14824; G553210; -.
DR   PIR; A01262; C4HU.
DR   PIR; A17265; A17265.
DR   PIR; A29177; A29177.
DR   PIR; B20807; B20807.
DR   HSSP; P01032; 1C5A.
DR   SWISS-2DPAGE; P01028; HUMAN.
DR   MIM; 120790; -.
DR   MIM; 120810; -.
DR   MIM; 120820; -.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN.
KW   COMPLEMENT PATHWAY; PLASMA; GLYCOPROTEIN; MHC III; SIGNAL;
KW   INFLAMMATORY RESPONSE; POLYMORPHISM; DISEASE MUTATION.
FT   SIGNAL        1     19
FT   CHAIN        20    675       BETA CHAIN.
FT   CHAIN       680   1443       ALPHA CHAIN.
FT   CHAIN      1451   1741       GAMMA CHAIN.
FT   PEPTIDE     680    756       C4A ANAPHYLATOXIN.
FT   DOMAIN      702    736       ANAPHYLATOXIN-LIKE.
FT   DISULFID    702    728       BY SIMILARITY.
FT   DISULFID    703    735       BY SIMILARITY.
FT   DISULFID    716    736       BY SIMILARITY.
FT   THIOLEST   1010   1013
FT   CARBOHYD    226    226       POTENTIAL.
FT   CARBOHYD    862    862       POTENTIAL.
FT   CARBOHYD   1328   1328       POTENTIAL.
FT   CARBOHYD   1391   1391       POTENTIAL.
FT   VARIANT     477    477       R -> W (IN C4A6).
FT   VARIANT     726    726       P -> L (IN C4A3).
FT   VARIANT    1073   1073       D -> G (IN C4A1, C4B1 AND C4B3).
FT   VARIANT    1120   1125       PCPVLD -> LSPVIH (IN C4B).
FT   VARIANT    1176   1176       N -> S (IN C4A1, C4B1, C4B3 AND C4B5).
FT   VARIANT    1201   1201       S -> T (IN C4A6, C4A3, C4A1, AND C4B).
FT   VARIANT    1207   1207       V -> A (IN C4A1, C4B1, C4B2 AND C4B3).
FT   VARIANT    1210   1210       L -> R (IN C4A1, C4B1, C4B2 AND C4B3).
FT   VARIANT    1286   1286       S -> A (IN C4A6, C4A1, C4A3A AND C4B).
FT   CONFLICT    727    727       D -> N (IN REF. 3).
FT   CONFLICT   1013   1013       Q -> E (IN REF. 4 AND 5).
SQ   SEQUENCE   1741 AA;  192335 MW;  405A8805 CRC32;

  Query Match            14.3%;  Score 1723;  DB 2;  Length 1741;
  Best Local Similarity  28.4%;  Pred. No. 0.00e+00;
  Matches  469;  Conservative  414;  Mismatches  622;  Indels  144;  Gaps  113;

Db     139 ghlflqtdqpiynpgqrvryrvfaldqkmrpst-dtitvmvenshglrvrkkevympssi 197

Qy     124 GFLFIHTDKPVYTPDQSVKVRVYSLNDDLKPAKRETVLTFID-PEGSEVDMVEEIDHIGI 182

Db     198 fq-ddfvipdisepgtwkisarfsdglesnsstqfevkkyvlpnfevkitpgkpyiltvp 256

Qy     183 ISFPDFKIPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEPEYNFI-GYK 241

Db     257 ghldemqldiqaryiygkpv-qgvayvrfglldе--dgkktffrglesqtklvnggshis 313

Qy     242 -NFKNFEITIKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLINGIAQVT 300

Db     314 lskaefqdaleklnm-gitdlqglrlyvaaaiiespggemeeaeltswyfvsspfsldls 372
```



```
Qy     301 FD-SE--TAVKELSYYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLNLV 357

Db     373 ktkrhlvpgapfllqalvremsgspasgipvkvsa-tvsspgsvpeaqdiqqntd-gsgq 430

Qy     358 ATPLFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDDGV 417

Db     431 vsipiiipqtiselqlsvsagsph-pai--arltvaapp-sg-gpgflsierpds-rppr 484

Qy     418 ASFVLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKALL 477

Db     485 vgdtlnlnlravgs-gatfshyyymilsrgqiv-fmnre--pkrtltsvsvfvdhhlaps 540

Qy     478 VGEHLNIIVTPKSPYIDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQNMVPS 537

Db     541 fyfvafyy-hgdh-p--vanslrvdvqagacegklelsvd-gakqyrngesvklhletds 595

Qy     538 SRLLVYYIVTGEQTAELVSDSVWLNIEE-KCGNQLQVHLSPDADAYSPGQTVSLNMATGM 596

Db     596 lalvalgaldtalyaagskshkplnmgkvfeamnsydlgcgpgggdsalqvfqaaglaf- 654

Qy     597 DSWVALAAVDSAVYGVQRGAKKPLE--RVFQFLEKSDLGCGAGGGLNNANVFHLAGLTFL 654

Db     655 sdgdqwtlsrkrlscpkekttrkkrnvnfqkainеklgqyasptakrccqdgvtrlpmmr 714

Qy     655 TNANADDSQENDEPC-KE-ILRPRRTL--QKKIEEIAAKYKHSVVKKCCYDG-ACVNNDE 709

Db     715 sceqraarvqq-pdcrepflsccqfaeslrkksrdkgqaglqraleilqeedlideddip 773

Qy     710 TCEQRAARISLGPRCIKAFTECCVVASQLR--A-NISHKDMQ--LGRLHMKTLLPVSKPE 764

Db     774 vrsffpenwlwrvetvdrfqiltlwlpdslttweihglslsktkglcvatpvqlrvfref 833

Qy     765 IRSYFPESWLWEVHLVPRRKQLQFALPDSLTTWEIQGIGISNT-GICVADTVKAKVFKDV 823

Db     834 hlhlrlpmsvrrfeqlelrpvlynyldknltvsvhvspveglclagggglaqqvlvpags 893

Qy     824 FLEMNIPYSVVRGEQIQLKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTKSSKC 883

Db     894 arp-vafsvvptaaaav-slkvvarg-sfefpv--g-davskvlqiekegaihreelvye 947

Qy     884 VRQKVEGSSSHLVTFTVLPLEIGLHNINFSLETWFGKEILVKTLRVVPEG-VKRES--YS 940

Db     948 lnpldhrgrtleipgnsd-pnmip-dgdfnsyv-rvtasdpldtlgse-ga-lspggvas 1002

Qy     941 GVTLDPRGIYGTISRRKEFPYRIPLDLVPKTEIKRILSVKGL-LVGEILSAVLSQEGINI 999

Db    1003 llrlprgcgeqtmiylaptlaasryldkteqwstlppet-kd-havdliqk-gymriqqf 1059

Qy    1000 LTHLPKGSAEAELMSVVPVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEGMLSIMSY 1059

Db    1060 rkadgsyaawlsrdsstwltafvlkvlslaqeqvggspeklqetsnwllsq-qqadgsfq 1118

Qy    1060 RNADYSYSVWKGGSASTWLTAFALRVLGQVNKYVEQNQNSICNSLLWLVENYQLDNGSFK 1119

Db    1119 dpcpvldrsmqggl-v-gndetvaltafvtialhhglavfqdegaeplkqrveasiskan 1176

Qy    1120 ENSQYQPIKLQGTLPVEARENSLYLTAFTVIGIRK--A-F-DI-C-PLV-KIDTALIKAD 1172

Db    1177 sflgekasagllgahaaaitayalslt-kapvdllgvahnnlmamaqetgdn-lywgsvt 1234

Qy    1173 NFLLENTLPAQ-STFTLAISAYALSLGDKTHPQFRSIV-SALKREALVKGNPPIYRFWKD 1230

Db    1235 gsqsnavsptpaprnpsdpmpqapalwiettayallhlllhegkaemadqasawltrqgs 1294
```

```
Qy   1231 NLQHKD-S-S-VP-N-TGT---AR-M-VETTAYALLTSL-NLKDINYVNPVIKWLSEEQR 1279

Db   1295 fqggfrstqdtvialdalsaywiashtteerglnvtlsstgrngfkshalqlnnrqirgl 1354
           : ||| |||||: |:::|: | :      : | |:: :   :   :| |    |:|   : :
Qy   1280 YGGGFYSTQDTINAIEGLTEYSLL--VKQLR-LSMDID-VS---YK-HKGALHN--YK-M 1328

Db   1355 eeelqfslgskinvkvggnskgtlkvlrtynvldmknttcqdlqievtvkghveytmean 1414
            :    | ||  ::| :  |     | |    :      :  :|   :::  ||  |   | |
Qy   1329 TDK-NF-LGRPVEVLL--ND-D-LIVSTGFGS-GL--AT---VHV-TTV-VHKTSTSEEV 1374

Db   1415 edyeydelpakddpdaplqpvtplqlfegrrnrrrreapkvveeqesrvhytvciwrngk 1474
             : |   |  | | :  ::: :  : : |  : :|       :|         |    : ::
Qy   1375 CSF-Y--L--KIDTQD-IEA-SHYRGY-GNSDYKR-----IVA---C-ASYKPS-REESS 1416

Db   1475 vglsgmaiadvtllsgfhalradlekltslsdryvshfetegphvllyfdsvptsr-ecv 1533
            | |: |: |::| :|: |     || |     |:  : ::     ||:| ::|:|:|    ||
Qy   1417 SG-SSHAVMDISLPTGISANEEDLKALVEGVDQLFTDYQIKDGHVILQLNSIPSSDFLCV 1475

Db   1534 gfeavqevpvglvqpasatlydyynperrcsvfygapsksrllatlcsaevcqcaegkcp 1593
            |   :   ||:: ||: |:|:|  |:::|::||:: |: ::    :| :  | | |: |
Qy   1476 RFRIFELFEVGFLSPATFTVYEYHRPDKQCTMFYST-SNIKI-QKVCEGAACKCVEADCG 1533

Db   1594 rqrralerglqdedgyrmkfacyyprveygfqvkvlredsraafrlfetkitqvlhftkd 1653
           : :  |:  :   :  |:  ||  | : |:: | :              |  : : : ::
Qy   1534 QMQEELDLTI-SAET-RKQTACK-PEIAYAYKVSITSITVENVFVKYKATLLDIYKTGEA 1590

Db   1654 vkaaanqmrnflvrasc-rlrlepgkeylimgldg--atydleghpqylldsnswieemp 1710
           |     :::  |: : :|     |  |::||||| ::     |::  :  | ||| :|||  |
Qy   1591 VAEKDSEI-TFIKKVTCTNAELVKGRQYLIMGKEALQIKYNFSFRYIYPLDSLTWIEYWP 1649

Db   1711 serlcrstrqraacaqlndflqeygtqgc 1739
            :  | |: | |   | |::| ::     ||
Qy   1650 RDTTC-SSCQ-AFLANLDEFAEDIFLNGC 1676
```

```
RESULT   12
ID   C04_MOUSE      STANDARD;      PRT;   1738 AA.
AC   P01029;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   COMPLEMENT C4 PRECURSOR (CONTAINS: C4A ANAPHYLATOXIN).
GN   C4.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85298264.
RA   SEPICH D.S., NOONAN D.J., OGATA R.T.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 82:5895-5899(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B12.WR;
RX   MEDLINE; 87309760.
RA   ROSA P.A., SEPICH D.S., ROBINS D.M., OGATA R.T.;
RL   J. IMMUNOL. 139:1568-1577(1987).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B12.WR; TISSUE=LIVER;
RX   MEDLINE; 89380278.
RA   OGATA R.T., ROSA P.A., ZEPF N.E.;
RL   J. BIOL. CHEM. 264:16565-16572(1989).
RN   [4]
```



```
RP   SEQUENCE FROM N.A.
RC   STRAIN=FM; TISSUE=LIVER;
RX   MEDLINE; 85289294.
RA   NONAKA M., NAKAYAMA K., YEUL Y.D., TAKAHASHI M.;
RL   J. BIOL. CHEM. 260:10936-10943(1985).
RN   [5]
RP   SEQUENCE OF 651-810 AND 924-1083 FROM N.A.
RX   MEDLINE; 85038607.
RA   NONAKA M., TAKAHASHI M., NATSUUME-SAKAI S., NONAKA M., TANAKA S.,
RA   SHIMIZU A., HONJO T.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 81:6822-6826(1984).
RN   [6]
RP   SEQUENCE OF 1105-1449 FROM N.A.
RX   MEDLINE; 85166208.
RA   LEVI-STRAUSS M., TOSI M., STEINMETZ M., KLEIN J., MEO T.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 82:1746-1750(1985).
RN   [7]
RP   SEQUENCE OF 1257-1376 FROM N.A.
RX   MEDLINE; 85038859.
RA   TOSI M., LEVI-STRAUSS M., DUPONCHEL C., MEO T.;
RL   PHILOS. TRANS. R. SOC. LOND., B, BIOL. SCI. 306:389-394(1984).
RN   [8]
RP   SEQUENCE OF 1360-1511 FROM N.A.
RX   MEDLINE; 83273751.
RA   OGATA R.T., SHREFFLER D.C., SEPICH D.S., LILLY S.P.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 80:5061-5065(1983).
RN   [9]
RP   SEQUENCE OF 1-128 FROM N.A.
RC   STRAIN=FM; TISSUE=LIVER;
RX   MEDLINE; 86031969.
RA   NONAKA M., NAKAYAMA K., YEUL Y.D., SHIMIZU A., TAKAHASHI M.;
RL   IMMUNOL. REV. 87:81-99(1985).
CC   -!- FUNCTION: C4 PLAYS A CENTRAL ROLE IN THE ACTIVATION OF THE
CC       CLASSICAL PATHWAY OF THE COMPLEMENT SYSTEM. IT IS PROCESSED BY
CC       ACTIVATED C1 WHICH REMOVE FROM THE ALPHA CHAIN THE C4A
CC       ANAPHYLATOXIN.
CC   -!- SUBUNIT: THIS PROTEIN IS SYNTHESIZED AS A SINGLE-CHAIN PRECURSOR
CC       AND, PRIOR TO SECRETION, IS ENZYMATICALLY CLEAVED TO FORM A TRIMER
CC       OF NONIDENTICAL CHAINS (ALPHA, BETA, AND GAMMA).
CC   -!- C4 IS A MAJOR HISTOCOMPATIBILITY COMPLEX CLASS-III PROTEIN.
CC   -!- SIMILARITY: TO C3, C5 AND ALPHA-2-MACROGLOBULIN.
CC   -!- SIMILARITY: CONTAINS ONE ANAPHYLATOXIN-LIKE DOMAIN.
DR   EMBL; K00019; E19386; -.
DR   EMBL; M11729; G387420; -.
DR   EMBL; M12968; G199270; -.
DR   EMBL; M12970; G199262; -.
DR   EMBL; M12972; G199264; -.
DR   EMBL; M11789; G387438; -.
DR   EMBL; K02798; G199282; -.
DR   EMBL; M17440; G387440; -.
DR   PIR; A01264; A01264.
DR   PIR; A21692; A21692.
DR   PIR; A22039; A22039.
DR   PIR; A24558; A24558.
DR   PIR; A29059; A29059.
DR   HSSP; P01032; 1C5A.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN.
KW   COMPLEMENT PATHWAY; PLASMA; GLYCOPROTEIN; MHC III; SIGNAL;
KW   INFLAMMATORY RESPONSE.
FT   SIGNAL        1     19
FT   CHAIN        20    673       BETA CHAIN.
FT   CHAIN       678   1443       ALPHA CHAIN.
FT   CHAIN      1448   1738       GAMMA CHAIN.
FT   PEPTIDE     678    753       C4A ANAPHYLATOXIN.
```

```
FT   DOMAIN      700    734       ANAPHYLATOXIN-LIKE.
FT   DISULFID    700    726       BY SIMILARITY.
FT   DISULFID    701    733       BY SIMILARITY.
FT   DISULFID    714    734       BY SIMILARITY.
FT   THIOLEST    1006   1009      BY SIMILARITY.
FT   CARBOHYD    224    224       POTENTIAL.
FT   CARBOHYD    743    743
FT   CARBOHYD    1387   1387      POTENTIAL.
FT   CONFLICT    132    132       F -> Y (IN REF. 4).
FT   CONFLICT    327    327       G -> E (IN REF. 4).
FT   CONFLICT    570    570       Q -> E (IN REF. 4).
FT   CONFLICT    720    720       R -> G (IN REF. 5).
FT   CONFLICT    739    740       DL -> AI (IN REF. 5).
FT   CONFLICT    838    838       P -> R (IN REF. 4).
FT   CONFLICT    993    993       P -> L (IN REF. 5).
FT   CONFLICT    1043   1043      D -> E (IN REF. 5).
FT   CONFLICT    1119   1119      V -> A (IN REF. 6).
FT   CONFLICT    1190   1190      A -> T (IN REF. 6).
FT   CONFLICT    1324   1324      K -> N (IN REF. 4).
FT   CONFLICT    1401   1401      G -> S (IN REF. 8).
FT   CONFLICT    1442   1442      R -> K (IN REF. 4).
FT   CONFLICT    1453   1453      A -> V (IN REF. 4).
SQ   SEQUENCE    1738 AA;  192870 MW;  A6CC1FCB CRC32;

  Query Match            13.3%;  Score 1605;  DB 2;  Length 1738;
  Best Local Similarity  27.6%;  Pred. No. 0.00e+00;
  Matches  449;  Conservative  407;  Mismatches 628;  Indels 143;  Gaps 111;

Db    137 ghifvqtdqpiynpgqrvryrvfaldqkmrpstdfltitvenshglrv-lkkei-ftsts 194
          | :|::|| |:| | | |: ||::|:: ::|:      :|  :: |  | :  ||    :
Qy    124 GFLFIHTDKPVYTPDQSVKVRVYSLNDDLKPAKRETVLTFIDPEGSEVDMVEEIDHIGII 183

Db    195 ifqdaftipdisepgtwkisarfsdglesnrsthfevkkyvlpnfevkitpwkpyilmvp 254
           | | | ||   :   | | | |:: : : :  :: |||| |||:| | | |    | ::
Qy    184 SFPD-FKIPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEP-E-YNFIGY 240

Db    255 snsdeiqldiqaryiygkpv-qgvaytrfalmde--qgkrtflrgletqaklvegrthis 311
            |  :::: | |||:| | | ::   |  |:: ::  :  :  ::     ::|::| ::::
Qy    241 KNFKNFEITIKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLINGIAQVT 300

Db    312 iskdqfqaaldkinigvrdleglrlyaatavlespggemeeaeltswrfvssafsldlsr 371
           :   :      |        :: ||:     || | | :|||| ||  ||||: : ::| |:: |:|
Qy    301 FDSETAVKELS--YYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLNLVA 358

Db    372 tkrhlvpgahfllqalvqemsgseasnvpvkvsa-tlvsgsdsqvldiqqstngiqq-v- 428
           |   | ||  : :     | :      :  ||| ::| |:     ::  ||    |   : : |
Qy    359 TPLFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDDGVA 418

Db    429 sisfpipptvtelrllvsagsl-ypai--arltvqa-p-psrgtgflsie-pldprspsv 482
           |: : :|: || | : | : :      |    ||     |: : :| : ::| |:    : :: |
Qy    419 SFVLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKALLV 478

Db    483 gdtfilnlqpvgipaptfshyyymiisrgqimamg-rep-rktv-tsvsvlvdhqlapsf 539
           |: : : : |  :       ::|| |:|:|:| |:  | ||     :    |::: | : : ||
Qy    479 GEHLNIIVTPKSPYIDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQNMVPSS 538

Db    540 yfvayfyhqg-h-p--vansllliniqsrdcegklqlkvd-gakeyrnadmmklriqtdsk 594
            :: |:   | : :   |::|: :||: : |    ||: :    |   |   :: : | : |
Qy    539 RLLVYYIVTGEQTAELVSDSVWLNIEEK-CGNQLQVHLSPDADAYSPGQTVSLNMATGMD 597

Db    595 alvalgavdtalyavggrshkpldmskvfevinsynvgcgpggddalqvfqdaglafsd 654
           : |||:|||:|:|:|  : |||:  :||: :: ::|||:||| :  ||: |||:|
Qy    598 SWVALAAVDSAVYGVQRGAKKPLE--RVFQFLEKSDLGCGAGGGLNNANVFHLAGLTFLT 655
```



```
Db    655 gdrltqtre-dlscpkekksrqkrnvnfqkavseklgqyssppdakrccqdgmtklpmkrt 713
              :::| | :| ||   | :| :  || : |  : |  :  |:|| ||   :    |
Qy    656 NANADDSQENDEPC-KEIL-RPRRTL--QKKIEEIAAKYKHSVVKKCCYDGAC-VNNDET 710

Db    714 ceqraarvpqqa-crepflscckfaedlrrnqtrsqahlarnnhnmlqeedlideddilv 772
          |||||||:: :  |  :|  ||  | :|| | :: : :|:|  | |     |:       :
Qy    711 CEQRAARISLGPRCIKAFTECCVVASQLRANISHKDMQLGRL-H-M-K--TLLPVSKPEI 765

Db    773 rtsfpenwlwrvepvdsskllltvwlpdsmttweihgvslskskglcvakptrvrvfrkfh 832
          |: |||:||| |   |    | |   ||||:|||||:|:::|:: |:|||   : :||:
Qy    766 RSYFPESWLWEVHLVPRRKQLQFALPDSLTTWEIQGIGISNT-GICVADTVKAKVFKDVF 824

Db    833 lhlrlpisirrfeqfelrpvlynylnddvavsvhvtpveglclagggmmaqqvtvpagsa 892
          | : :| |: | ||::|:  :|||     :      | :::|||:| : : :: :| | ::
Qy    825 LEMNIPYSVVRGEQIQLKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTKSSKCV 884

Db    893 rp-vafsvvptaaanv-plkvvargv-fdl----g-davskilqiekegaihreelv-yn 943
          |  |  |     :  | || :  : : | |     | : : | |::  || : ||
Qy    885 RQKVEGSSSHLVTFTVLPLEIGLHNINFSLETWFGKEILVKTLRVVPEG-VKRESYSGVT 943

Db    944 ldp--l-nnlgrtleipgssdpnivpdgdfsslvrvtasepletmgsegalspggvasll 1000
          |||  :   ::|  |:|       ::||  ::  :: | :     | : | : || |:  |
Qy    944 LDPRGIYGTISRRKEFPYRIPLDLVPKTEIKRILSVKGLLVGEIL-S-AVLSQEGINILT 1001

Db   1001 rlpqgcaeqtmiylaptltasnyldrteqwsklspet-kd-havdliqk-gymriqqfrk 1057
          :|| | ||   :: : | :    :||:  ::|: : ::    : : :      | | :  |  :|:
Qy   1002 HLPKGSAEAELMSVVPVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEGMLSIMSYRN 1061

Db   1058 ndgsfgawlhrdsstwltafvlkilslaqeqvgnspeklqetaswllaq-qlgdgsfhdp 1116
            | |:: |     :||||||| |::|:      |   : : : : ::   ||:     || :||| :
Qy   1062 ADYSYSVWKGGSASTWLTAFALRVLGQVNKYVEQNQNSICNSLLWLVENYQLDNGSFKEN 1121

Db   1117 cpvihramqggl-vgs-detvaltafvvialhhgldvfqdddakqlknrveasitkansf 1174
                :|| | | : :::: |||| ||::: :    |   |        | :::::: ||::|
Qy   1122 SQYQPIKLQGTLPVEARENSLYLTAFTVIGIRKA---F--DIC-PL-VKIDTALIKADNF 1174

Db   1175 lgqkasagllgahaaaitayaltlt-kasedlrnvahnslmamaeetgeh-lywglvlgs 1232
          | ::: ::   :: : ||:|||:|  |: ::|::  ::|:   |    |:  :|:    :
Qy   1175 LLENTLPAQ-STFTLAISAYALSLGDKTHPQFRSIV-SALKREALVKGNPPIYR---F-W 1228

Db   1233 qdkvvlrptaprsptepvpqapalwiettayall-hlllregkgkmadkaaswlthqgsf 1291
            |:  |:    :   | :     |  : :|||||||  | |::    :   :      ||: :  :
Qy   1229 KDN--LQHKDSSVP-NTGT-AR-M-VETTAYALLTSLNLKD-I-NYVNPVIKWLSEEQRY 1280

Db   1292 hgafrstqtdtvvtldalsaywiashtteekalkvtlssmgrnglkthglhlnnhqvkgle 1351
          |:| |||||: ::::|:  |    |   | |::::  :  | |   |:|  |:|   | :
Qy   1281 GGGFYSTQDTINAIEGLTEY---SLLV--KQLRLSMD-ID-VSYKHKGA-LHNY--K-MT 1329

Db   1352 eelkfslgstisvkvegnskgtlkilrtynvldmknttcqdlqievkvtgaveyawdane 1411
           :  :| || : | :  |     |  | |: |     :    |  |  ||   | :  ::|
Qy   1330 DK-NF-LGRPVEVLL--ND-D-L-IVST-G-FG----S--GLAT-VHVTTVVHKT-STSE 1372

Db   1412 dyedyydmpaaddpsvplqpvtplqlfegrrsrrrreapkvaeeqesrvqytvciwrngk 1471
            :  :|  :  |   :  ::  :   :   :  |  | :|    ||      |      : ::
Qy   1373 EVCSFY-LKI-DTQDI--EA-SHYRGY-GN-SDYKR-I--VA----C-ASYKPS-REESS 1416

Db   1472 lglsgmaiaditllsgfhalradlekltslsdryvshfetdgphvllyfdsvpttr-ecv 1530
           | |: |: ||:| :|: |   ||   |   |:  :  ::    ||:| ::|:|::   ||
Qy   1417 SG-SSHAVMDISLPTGISANEEDLKALVEGVDQLFTDYQIKDGHVILQLNSIPSSDFLCV 1475

Db   1531 gfgasqevvvglvqpssavlydyyspdhkcsvfyaaptksqllatlcsgdvcqcaegkcp 1590
            |   :    ||:: |::  :|:| ||  |::||:: ::  :   :| |  | | | |: |
Qy   1476 RFRIFELFEVGFLSPATFTVYEYHRPDKQCTMFYST-SNIKI-QKVCEGAACKCVEADCG 1533
```

```
Db   1591 rllrslerrvedkdgyrmrfacyyprveygftvkvlredgraafrlfeskitqvlhfrkd 1650
           ::   |:  :    :  |:: ||  | : |:: | :         |  : : : ::     :
Qy   1534 QMQEELDLTISA-ET-RKQTACK-PEIAYAYKVSITSITVENVFVKYKATLLDIYK-TGE 1589

Db   1651 tmasigqtrnflsrasc-rlrlepnkeylimgmdgetsdnkgdpqyl--ldsntwieemp 1707
           ::|       |: : :|    |    ::||||:::     :    :|:  ||| |||| |
Qy   1590 AVAEKDSEITFIKKVTCTNAELVKGRQYLIMGKEALQIKYNFSFRYIYPLDSLTWIEYWP 1649

Db   1708 seqmcks 1714
            :  | |
Qy   1650 RDTTCSS 1656


RESULT   13
ID   CO3_RABIT      STANDARD;      PRT;   726 AA.
AC   P12247;
DT   01-OCT-1989 (REL. 12, CREATED)
DT   01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   COMPLEMENT C3 ALPHA CHAIN (FRAGMENT).
GN   C3.
OS   ORYCTOLAGUS CUNICULUS (RABBIT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; LAGOMORPHA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 87006907.
RA   KUSANO M., CHOI N.H., TOMITA M., YAMAMOTO K., MIGITA S., SEKIYA T.,
RA   NISHIMURA S.;
RL   IMMUNOL. INVEST. 15:365-378(1986).
CC   -!- FUNCTION: C3 PLAYS A CENTRAL ROLE IN THE ACTIVATION OF THE
CC       COMPLEMENT SYSTEM. ITS PROCESSING BY C3 CONVERTASE IS THE CENTRAL
CC       REACTION IN BOTH CLASSICAL AND ALTERNATIVE COMPLEMENT PATHWAYS.
CC       AFTER ACTIVATION C3B CAN BIND COVALENTLY, VIA ITS REACTIVE
CC       THIOLESTER, TO CELL SURFACE CARBOHYDRATES OR IMMUNE AGGREGATES.
CC   -!- SUBUNIT: C3 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 ARG
CC       RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE
CC       BOND. C3 CONVERTASE ACTIVATES C3 BY CLEAVING THE ALPHA CHAIN,
CC       RELEASING C3A ANAPHYLATOXIN & GENERATING C3B (BETA CHAIN + ALPHA'
CC       CHAIN).
CC   -!- SIMILARITY: TO C4, C5 AND ALPHA-2-MACROGLOBULIN.
DR   EMBL; M32434; G164863; -.
DR   PIR; A27602; A27602.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN.
KW   COMPLEMENT PATHWAY; COMPLEMENT ALTERNATE PATHWAY; PLASMA;
KW   INFLAMMATORY RESPONSE; GLYCOPROTEIN.
FT   NON_TER       1      1
FT   CHAIN        <1    726       ALPHA CHAIN.
FT   THIOLEST     73     76
FT   CARBOHYD      2      2       POTENTIAL.
FT   CARBOHYD    233    233       POTENTIAL.
FT   CARBOHYD    680    680       POTENTIAL.
SQ   SEQUENCE   726 AA;  81844 MW;  27A509CF CRC32;

  Query Match             6.6%;  Score 796;  DB 2;  Length 726;
  Best Local Similarity   26.2%;  Pred. No. 5.45e-148;
  Matches   197;  Conservative  197;  Mismatches  305;  Indels   52;  Gaps   37;

Db      9 tldpenlgqggv-qkeeipsadisdqvpgtesetkil-lqgtpvaqmtedaidgerlkhl 66
          ||||  :  | : :: |:|     | || ||   :|| : |  |:::     :  | :: |
Qy    943 TLDPRGI-YGTISRRKEFPYRIPLDLVPKTEI-KRILSVKGLLVGEILSAVLSQEGINIL 1000

Db     67 ivtgsgcgeqnmiamthtviavhyldhteqwdkf-s--lekrqealelikkgytqqlafk 123
              | :| :::::      |||:  ::|: | |  |  :|     :| |     ::::
```



```
Qy   1001 THLPKGSAEAELMSVVPVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEGMLSIMSYR 1060

Db    124 qpnsayaaflnrapstwltayvvkvfslavnliaidsqvlcgavkwlimekqkpdgvfqe 183
            :: :|:      ::||||||: ::|::    : :   :   :| :: ||: : |   :| | |
Qy   1061 NADYSYSVWKGGSASTWLTAFALRVLGQVNKYVEQNQNSICNSLLWLVENYQLDNGSFKE 1120

Db    184 dapvihqemigggr-nsvekeraltafvlialqeareiceeqvnslaasinksrdflaan 242
          ::       : |     :: |:    |||| :|::: | :||   |: : ::: |: :|| |
Qy   1121 NSQYQPIKLQGTLPVEAREN SLYLTAFTVIGIRKAFDICP-LVK-IDTALIKADNFLLEN 1178

Db    243 ymnlqrpysvaiaayawaqqdk----lrg--aflnk-flskakek-nr-weepgq-r--- 289
            :  |  :::||:||| :   ||     :|:  : |::  | |::      | | :   | :
Qy   1179 TLPAQSTFTLAISAYALSLGDKTHPQFRSIVSALKREALVKGNPPIYRFWKDNLQHKDSS 1238

Db    290 l-yn-----veassyallalllrdfdsvppvvrwlneqryygggygstqatfmgfqala 343
            :          ||:::||||: | |:|:: | ||::||:|:: ||||: ||| |: ::::|:
Qy   1239 VPNTGTARMVETTAYALLTSLNLKDINYVNPVIKWLSEEQRYGGGFYSTQDTINAIEGLT 1298

Db    344 qyqtdvpdhkdlnmvvsiqlpsrsspvkhrivwdsasllrseetkenqgfsltaq-gkgq 402
            :|   | : : |:| : :     ::: : ::   |     | |: |     |: : :::   | |
Qy   1299 EYSLLVKQLR-LSMDIDVSYKHKGALHNYKMT-DKNFLGRPVEVLLNDDLIVSTGFGSGL 1356

Db    403 gtlsvvttyfakvkgkvtckkfdlrvniktapetvkkpqdakstmilghctry-lg-ded 460
          :|: | |         :    |  | |:::        :  :          : :  |: |  : :|
Qy   1357 ATVHVTTVVHKTSTSEEVCS-FYLKIDTQDIEASHYRGYGNSDYKRIVACASYKPSREES 1415

Db    461 at-ms--ildismmtgfvpdtddlnllstgvdryiskyelnkafsnkntliiyldkishs 517
          ::   |   ::|||: ||: :: :||: |    |||: :: |::  |   ::    :|: |: |: |
Qy   1416 SSGSSHAVMDISLPTGISANEEDLKALVEGVDQLFTDYQI-KD-GH---VILQLNSIPSS 1470

Db    518 reeclafkvhqyfnvgliqpgavkvysyynleetctqfyhpekedgmlsklchkemcrca 577
             |: |:: : |:||:: |::   || |     :  || ||  :   : :: |:|      |:|
Qy   1471 DFLCVRFRIFELFEVGFLSPATFTVYEYHRPDKQCTMFY-S-TSNIKIQKVCEGAACKCV 1528

Db    578 eencf-mqq-ldekitlndrldkacepgldyvyktklvqveraddfdeylmvventiksg 635
            | :|  ||: ||  |: : : : || | : | ||  :  :    : |   |   : :  |:|
Qy   1529 EADCGQMQEELDLTISAETRKQTACKPEIAYAYKVSITSITVENVFVKYKATLLDIYKTG 1588

Db    636 sdevqagqpapfishikcrdalklkdgkhylmwglssdpvgekpntsyiigkd--twvef 693
                :         ||  :  | :|  |  |::||: |  :   :  : :  ||    |  ||:|:
Qy   1589 EAVAEKDSEITFIKKVTCTNA-ELVKGRQYLIMGKEALQIKYNFSFRYIYPLDSLTWIEY 1647

Db    694 wpekeecqdeenqkhcedlgafaesmvvfgc 724
            ||    |      |      :|  ||| : : ||
Qy   1648 WPRDTTC-SS-CQAFLANLDEFAEDIFLNGC 1676


RESULT   14
ID   A1I3_RAT       STANDARD;      PRT;  1477 AA.
AC   P14046;
DT   01-APR-1990 (REL. 14, CREATED)
DT   01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   ALPHA-1-INHIBITOR III PRECURSOR.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=LIVER;
RX   MEDLINE; 88153707.
RA   BRACIAK T.A., NORTHEMANN W., HUDSON G.O., SHIELS B.R., GEHRING M.R.,
RA   FEY G.H.;
RL   J. BIOL. CHEM. 263:3999-4012(1988).
```

```
CC   -!- FUNCTION: PROTEASE INHIBITOR WITH A WIDE SPECTRUM OF PROTEIN
CC       TARGETS, WHICH ATTACHES THROUGH ITS THIOLESTER FUNCTION.
CC   -!- THIS INHIBITOR HAS REDUCED PLASMA CONCENTRATIONS IN ACUTE PHASE.
CC   -!- SUBUNIT: MONOMER.
CC   -!- SIMILARITY: TO OTHER PROTEINS OF THE ALPHA-MACROGLOBULIN FAMILY,
CC       INCLUDING COMPLEMENT COMPONENTS C3, C4, AND C5.
DR   EMBL; J03552; G202572; -.
DR   EMBL; M22358; G554401; -.
DR   PIR; A29952; A29952.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN.
KW   SERINE PROTEASE INHIBITOR; GLYCOPROTEIN; PLASMA; BAIT REGION; SIGNAL;
KW   LIVER.
FT   SIGNAL        1     24
FT   CHAIN        25   1477       ALPHA-1-INHIBITOR III.
FT   DOMAIN      601    750       BAIT REGION (APPROXIMATELY).
FT   DISULFID     48     86       BY SIMILARITY.
FT   DISULFID    251    295       BY SIMILARITY.
FT   DISULFID    269    283       BY SIMILARITY.
FT   DISULFID    468    563       BY SIMILARITY.
FT   DISULFID    595    774       BY SIMILARITY.
FT   DISULFID    643    678       BY SIMILARITY.
FT   DISULFID    850    886       BY SIMILARITY.
FT   DISULFID    924   1324       BY SIMILARITY.
FT   DISULFID   1082   1130       BY SIMILARITY.
FT   DISULFID   1355   1470       BY SIMILARITY.
FT   THIOLEST    975    978
FT   CARBOHYD     55     55       POTENTIAL.
FT   CARBOHYD    247    247       POTENTIAL.
FT   CARBOHYD    301    301       POTENTIAL.
FT   CARBOHYD    321    321       POTENTIAL.
FT   CARBOHYD    393    393       POTENTIAL.
FT   CARBOHYD    508    508       POTENTIAL.
FT   CARBOHYD    750    750       POTENTIAL.
FT   CARBOHYD    777    777       POTENTIAL.
FT   CARBOHYD    872    872       POTENTIAL.
FT   CARBOHYD    994    994       POTENTIAL.
FT   CARBOHYD   1143   1143       POTENTIAL.
FT   CARBOHYD   1314   1314       POTENTIAL.
FT   CARBOHYD   1427   1427       POTENTIAL.
FT   VARIANT       4      4       D -> N.
FT   VARIANT      87     87       V -> L.
FT   VARIANT     137    137       M -> V.
SQ   SEQUENCE   1477 AA;  163773 MW;  8FAE9F1B CRC32;

  Query Match             5.6%;  Score 670;  DB 1;  Length 1477;
  Best Local Similarity   23.6%;  Pred. No. 2.99e-118;
  Matches  365;  Conservative  419;  Mismatches  615;  Indels  146;  Gaps  121;

Db       14 allaflpfasllngnsk-ymvlvpsqlytetpekiclhlyhlnetvtvtaslisqrgtrk 72
            ::| || | :    |     |:: |  :  ::|:| :::|  |:   | |: |      :
Qy        5 GILCFLIFLGKTWGQEQTYVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSYPDKKF 64

Db       73 lfde-lv-vdkdl-fh-cvsfti-pr-lpseeeesldiniegakhkfserrvvlvknke 126
              :      |  :  :   |:    :|| |: ||:::: |   : :|      ||  : : :   :
Qy       65 SYSSGHVHLSSENKFQNSAILTIQPKQLPGGQNPVSY-VYLEVVSKHFSKSKRMPITYDN 123

Db      127 svvfvqtdkpmykpqqsvkfrvvsmdknlhplnelfp.ayiedpkmnri-mqwqdvkten 185
            : :|::||||:| | |||| || |:: :| | :    |::| ||   : : |   :::
Qy      124 GFLFIHTDKPVYTPDQSVKVRVYSLNDDLKPAKRETVLTFI-DPEGSEVDMV-EEID-HI 180

Db      186 glkqls-fslsaepiqgpyki-vilkqs-gvkeehsftvmefvlprfgvdvkvpnaisvy 242
            |:  :: | ::::|  |    |    |:  :       | |:|:|||:|:| :       :  |
Qy      181 GIISFPDFKIPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEPEYNFIGY 240
```



```
Db      243 deiin--vtacatytygkpvp-ghvkisl-chgnptfssetksgckeedsrldnngcstq 298
              :: |   :|  | | | | |  : | |::  : :          :      ::: |  ||  :|
Qy      241 KNFKNFEITIKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTML-ING-IAQ 298

Db      299 evnitefqlke-ny--lk-mhqaf-hvnatvteegtgsefsgsgriever-trnkf-lfl 351
                :|  :|| :|  |  ::  :   :  || | :||: ||  : |   :    : : | |
Qy      299 VTFDSETAVKELSYYSLEDLNNKYLYIAVTVIE-STGG-FSEEAEIPGIKYVLSPYKLNL 356

Db      352 kadshf-rhgipffvkvrlvdikgdpipneqvlikardagytnattdqhglakfsidtn 410
              | : | : |||: :||:: |    : :      |:::|:          |:  : : :    :| :
Qy      357 VATPLFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVD-D 415

Db      411 qisdyslnikvyhkeesscihssctaerhaeah-htay-av-ys-lsksyiyld-tea-g 464
            |:: : ||:               : :   :::   | : : :| |: || || ||:|:| |:
Qy      416 GVASFVLNLPSGVTVLEFNVKTD-APDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHK 474

Db      465 vlpcnqihtvqahfilkgqvlgvlqqivfhylvmaqgsilqtgnhthqvepgesqvqgnf 524
              |    : | :          |:  :  : :  : ::||::: | |:: |  |::   :::  |
Qy      475 ALLVGE-H-LNIIVTPKSPYIDKITH--YNYLILSKGKIIHFG--TRE-KFSDASYQ--- 524

Db      525 aleipvefsmvpvakmliytilp-d--geviadsvkfqve-kclrn-kvhlsfspsqslp 579
            :::|||  :|||  :::|:| |:  :  :|:::|||  : :| ||       |||| : ::: :
Qy      525 SINIPVTQNMVPSSRLLVYYIVTGEQTAELVSDSVWLNIEEKCGNQLQVHLS-PDADAYS 583

Db      580 asqt-hmrvtaspqslcglravdqsvl-lq----kp-e-a-e-lsps-liydlpg-mqds 627
            ::||  : :::: :|  :| ||| :|  :|      || |    : |  |  |       | : ::
Qy      584 PGQTVSLNMATGMDSWVALAAVDSAVYGVQRGAKKPLERVFQFLEKSDLGCGAGGGLNNA 643

Db      628 n-f-ias-s--ndpf-ed-edyclmyqpiarek-dvyryvretglmaftnlkiklpty-- 677
            | | :|: :   ::  :| ::           | | :  : : : | : : : : :|    |
Qy      644 NVFHLAGLTFLTNANADDSQENDEPCKEILRPRRTLQKKIEEIAAK-YKHSVVKKCCYDG 702

Db      678 -cntdydmvplavpavaldsstdrgmyeslpvva-vksplpqepprkdppp-kd--pvie 732
              |  : :        : ::| :    ::   |    | : ::: :::    :         |   ||
Qy      703 ACVNNDETCEQRAARISLGPRCIKAFTECCVVASQLRANISHKDMQLGRLHMKTLLPVSK 762

Db      733 t-irnyfpetwiwdlvtvnssgvtelemtvpdtitewkagalclsndtglglssvasfqa 791
              ||:||||:|:|: | : :     :|: ::||::|  |     :: :|| ||: ::
Qy      763 PEIRSYFPESWLWE-VHLVPR-RKQLQFALPDSLTTWEIQGIGISN-TGICVADTVKAKV 819

Db      792 fqpffveltmpysvirgeaftlkatvlnylptslpmavlleaspdf-ta-vpvennqd-- 847
            |    |:|: :||||:|||  : ||:|| ||   :::     | : |   : |:  || ::|
Qy      820 FKDVFLEMNIPYSVVRGEQIQLKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTK 879

Db      848 sy-clqangrhtsswlvtpkslg-nvnf-svs-a-earqspgpcgsevatvpetgrkdtv 902
            |  |:   :   :|| |||    |   :: : ::: : |:: :          :   |||  ::::
Qy      880 SSKCVRQKVEGSSSHLVTFTVLPLEIGLHNINFSLETWFGKEILVKTLRVVPEGVKRESY 939

Db      903 vkvlivepegikkehtfssllcasdaelsetlslllpptvvkdsarahfsvmgdilssai 962
              |   ::| ||       ::::       :         | |:  | |  :|        ::|:|||: :
Qy      940 SCV-TLDPRGI-----YGTISRRKEFPYRIPLDLV-PKTEIKRILSVKGLLVGEILSAVL 992

Db      963 -kntqn-liqmpygcgeqnmvlfapniyvlkyln-etq-ql--tekikska-lqy-lrag 1014
              :   | ::| | :| ::::    | :|: ||:   :   :   :: :   |   |   |: |
Qy      993 SQEGINILTHLPKGSAEAELMSVVPVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEG 1052

Db     1015 yqrelnykhkdgsyafgdhngqgqgntwltafvlksfaqarafifideshitdaftwls 1074
            ::|:: |    || ::   :| |  : ::||||||  |: ::|     ::   ::: | ::: ||
Qy     1053 MLSIMSYRNAD--YS-YSVWKG-GSASTWLTAFALRVLGQVNKYVEQNQNSICNSLLWLV 1108

Db     1075 kq-qkdsgcfrssgsllnnamkggvddeitlsa-yit-malless1p-dtdpvvskalsc 1130
               | |:| |: ::          : | :   |    :: |:|  :::      : |   |:|        :
Qy     1109 ENYQLDNGSFKENSQYQPIKLQGTLPVEARENSLYLTAFTVIGIRKAFDICPLVKIDTAL 1168
```

```
Db   1131 lesswenieqggngsfvytkalmayafalagnqekr-neilksldkea-ikednsihwer 1188
          : :     :|     :     :| |: |||::|: :     :     |: :| :|| :| : :| : :
Qy   1169 IKADNFLLENTLPAQSTFTLAISAYALSLGDKTHPQFRSIVSALKREALVKGNPPIYRFW 1228

Db   1189 pqkptksegylytpqassaevemsayvvlarltaqpapspedlalsmgtikwltkqqnsy 1248
          ::     :: :     :::  || :|| |  ||:     :  |:        ||||: :|  |
Qy   1229 KDNLQHKDSSV-PNTGTARMVETTAYA-L--LTSL---NLKDINYVNPVIKWLSEEQR-Y 1280

Db   1249 gg-fsstqdtvvaldalskyga-atfsksqktpsvtvqssgsfsqkfqvdksnrlllqqv 1306
          || | |||||: |:::|: |:          :  :    |:     |:: ::: :    | :| : |
Qy   1281 GGGFYSTQDTINAIEGLTEYSLLVKQLRLSMDIDVSYKHKGAL-HNYKMTDKN-FLGRPV 1338

Db   1307 slpyipgnytvsvs-gegcvyaq-ttlrynvplekqqpafalkvqtvpltcnnpkg-qns 1363
           :  :  :  ||:'||    : ||:  : :    :   :| ||::|  :  ::  :|   ||
Qy   1339 EV-LLNDDLIVSTGFGSGLATVHVTTVVHKTSTSEEVCSFYLKIDTQDIEASHYRGYGNS 1397

Db   1364 -fqisle-isymgsr--pas-nm-viadvkmlsgfiplkptvkkl-erlghv-srtevtt 1415
            :   :   ||: ||   ::| :    : |: : :|: :     :| | | : :: :   ::
Qy   1398 DYKRIVACASYKPSREESSSGSSHAVMDISLPTGISANEEDLKALVEGVDQLFTDYQIKD 1457

Db   1416 nnvllyldqv-tnqtlsfsfiiqqdipvknlqpaivkvydyyetd 1459
           :|:| |: : ::: |    | | : : |   | ||     ||:|     |
Qy   1458 GHVILQLNSIPSSDFLCVRFRIFELFEVGFLSPATFTVYEYHRPD 1502
```

```
RESULT   15
ID   A2MG_MOUSE     STANDARD;      PRT;  1476 AA.
AC   P28665;
DT   01-DEC-1992 (REL. 24, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   MURINOGLOBULIN 1 PRECURSOR (MUG1).
GN   MUG1 OR MUG-1.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 28-57.
RC   TISSUE=LIVER;
RX   MEDLINE; 91358495.
RA   OVERBERGH L., TORREKENS S., VAN LEUVEN F., VAN DEN BERGHE H.;
RL   J. BIOL. CHEM. 266:16903-16910(1991).
CC   -!- FUNCTION: A PROTEINASE ACTIVATES THE INHIBITOR BY SPECIFIC
CC       PROTEOLYSIS IN THE BAIT REGION, WHICH, BY AN UNKNOWN MECHANISM
CC       LEADS TO REACTION AT THE CYSTEINYL-GLUTAMYL INTERNAL THIOL ESTER
CC       SITE AND TO A CONFORMATIONAL CHANGE, WHEREBY THE PROTEINASE IS
CC       TRAPPED AND/OR COVALENTLY BOUND TO THE INHIBITOR. WHILE IN THE
CC       TETRAMERIC PROTEINASE INHIBITORS STERIC INHIBITION IS SUFFICIENTLY
CC       STRONG, MONOMERIC FORMS NEED A COVALENT LINKAGE BETWEEN THE
CC       ACTIVATED GLUTAMYL RESIDUE OF THE ORIGINAL THIOL ESTER AND A
CC       TERMINAL AMINO GROUP OF A LYSINE OR ANOTHER NUCLEOPHILIC GROUP ON
CC       THE PROTEINASE, FOR INHIBITION TO BE EFFECTIVE.
CC   -!- TISSUE SPECIFICITY: PLASMA.
CC   -!- SUBUNIT: MONOMER.
CC   -!- SIMILARITY: TO OTHER PROTEINS OF THE ALPHA-MACROGLOBULIN FAMILY,
CC       INCLUDING COMPLEMENT COMPONENTS C3, C4, AND C5.
DR   EMBL; M65736; G199891; -.
DR   PIR; A41185; A41185.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN.
KW   SERINE PROTEASE INHIBITOR; GLYCOPROTEIN; PLASMA; BAIT REGION; SIGNAL;
KW   MULTIGENE FAMILY.
FT   SIGNAL         1     27
FT   CHAIN         28   1476       MURINOGLOBULIN 1.
FT   DOMAIN       677    734       BAIT REGION (APPROXIMATELY).
```



```
FT   DISULFID     48     86       BY SIMILARITY.
FT   DISULFID    251    276       BY SIMILARITY.
FT   DISULFID    269    288       BY SIMILARITY.
FT   DISULFID    461    555       BY SIMILARITY.
FT   DISULFID    587    773       BY SIMILARITY.
FT   DISULFID    634    680       BY SIMILARITY.
FT   DISULFID    849    885       BY SIMILARITY.
FT   DISULFID    923   1323       BY SIMILARITY.
FT   DISULFID   1081   1129       BY SIMILARITY.
FT   DISULFID   1354   1469       BY SIMILARITY.
FT   THIOLEST    974    977       BY SIMILARITY.
FT   CARBOHYD     55     55       POTENTIAL.
FT   CARBOHYD    294    294       POTENTIAL.
FT   CARBOHYD    313    313       POTENTIAL.
FT   CARBOHYD    500    500       POTENTIAL.
FT   CARBOHYD    749    749       POTENTIAL.
FT   CARBOHYD    776    776       POTENTIAL.
FT   CARBOHYD    871    871       POTENTIAL.
FT   CARBOHYD    993    993       POTENTIAL.
FT   CARBOHYD   1142   1142       POTENTIAL.
FT   CARBOHYD   1180   1180       POTENTIAL.
FT   CARBOHYD   1426   1426       POTENTIAL.
SQ   SEQUENCE   1476 AA;  165139 MW;  7E3EAF00 CRC32;

  Query Match             5.5%;  Score 661;  DB 1;  Length 1476;
  Best Local Similarity   24.1%;  Pred. No. 3.83e-116;
  Matches  368;  Conservative  407;  Mismatches  612;  Indels  140;  Gaps  112;

Db     28 dskymvlvpsqlytetpekiclhlyqlnetvtvtaslvsqsgrknlfde-lv-ldkdl-f 84
          :  |::  |  :    ::|:| :::|    |:    | |: | : :|  :      | |  :  |
Qy     20 EQTYVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSYPDKKFSYSSGHVHLSSENKF 79

Db     85 qcvsfi-i-pr-lsssdee-dflyvdikgpthefskrkavlvknkesvvfvqtdkpvykp 140
          |   ::: | |: |:::::   ::|:::  : | ||| | : :    :: :|::|||||| |
Qy     80 QNSAILTIQPKQLPGGQNPVSYVYLEV-VSKH-FSKSKRMPITYDNGFLFIHTDKPVYTP 137

Db    141 gqsvkfrvvsmdkmlrplnellplayiedpkknri-mqwrdiktenglkqmsfslaaepi 199
          |||| || |::  |:| :    |::| ||   : : |    :|       :      | ::::|
Qy    138 DQSVKVRVYSLNDDLKPAKRETVLTFI-DPEGSEVDMV-EEIDHIGIISFPDFKIPSNPR 195

Db    200 qgpykivv-hkesgekeehs-ftvmefvlprfnvdlkvp-namsvnd-evlsvtacgkyt 255
          |    |     ||        : | |:|:|||:|:| :     | :: ::    : :|   ::|
Qy    196 YGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEPEYNFIGYKNFKNFEITIKARYF 255

Db    256 ygkpvp-ghvkinv-cret-etgcrev-nsqldnng-cs-tqevni-telqskk-rnyev 307
          | | |   : | |      ||          :|:    : : ::|     :   :| : :|    |      | :
Qy    256 YNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLINGIAQVTFDSETAVKELSYYSL 315

Db    308 q-l---f-hvnatvteegtgl-efsrsgttkieritnkliflkadshf-rhgipffvkvr 360
          : |    :  :  || |    |: | :        |          || | | : | : |||: :||:
Qy    316 EDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLN-LVATPLFLKPGIPYPIKVQ 374

Db    361 lvdikgdpipnekvfikaqelsytsatttdqhglaefsidttcisgsslhikvnhkeeds 420
          : |    : :     |:::||  :      |:   : : :   :|    ::: |::          :
Qy    375 VKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDDG-VASFVLNLPSGVTVLEF 433

Db    421 csyfycmeerhasakhvay-av-ys-lsksyiyld-tetssilpcnqihtvqahfilkgd 476
                :    :  : :| |: || || ||:|:| |:     |    : | :        |:
Qy    434 NVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKALLVGE-H-LNIIVTPKSP 491

Db    477 lgvlkelifyylvmaqgsiiqtgnhthqvepgeapvkgkfaleipvefsmvpmakmliyt 536
             : |    : ||::: | ||: |   |::    ::|:       :::|||  :||| :::|:|
Qy    492 Y-IDKITHYNYLILSKGKIIHFG--TRE-KFSDASYQ---SINIPVTQNMVPSSRLLVYY 544
```

```
Db    537 ilp-d--geviadsvnfeie-kclrnkvdlrfst-sqslpasqt-rlqvtaspqslcglr 590
          |:  :  :|:::|||  ::|| ||  | :::::|  ::: :::||  | :::: :|  :|
Qy    545 IVTGEQTAELVSDSVWLNIEEKC-GNQLQVHLSPDADAYSPGQTVSLNMATGMDSWVALA 603

Db    591 avdqsvlllkpeselspswiynlpgmqqnkfvpssrlsedqedcilysswlaekhtnlvp 650
          ||| :|   :    :   :   ::: :   : :        ::: |::     :   : |:: :::
Qy    604 AVDSAVYGVQRGAKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLAGLTFLTNANADDSQ 663

Db    651 hgte--kdvyr---yved-mg-lta-ftnlmik---lpiicfdygmv-pisaprvefd-- 696
             |  |:: |     ::  :  ::| : : ::|         | :         |:|: :
Qy    664 ENDEPCKEILRPRRTLQKKIEEIAAKYKHSVVKKCCYDGACVNNDETCEQRAARISLGPR 723

Db    697 -l-aftp--eisws1rttlskrpeepprkdpssndpltet-irkyfpetwvwdivtvnst 751
           : |||     ::   ||: :| :  :   |       :  |::   || ||||:|:|::  |
Qy    724 CIKAFTECCVVASQLRANISHKDMQLGRLHMKTLLPVSKPEIRSYFPESWLWEVHLVPRR 783

Db    752 glaevemtvpdtitewkagalclsndtglglssvvplqafkpffvevslpysvvrgeafm 811
              ::: ::||::|  |    :: :|| ||: ::   |      ||   |:|:::|||||||| :
Qy    784 --KQLQFALPDSLTTWEIQGIGISN-TGICVADTVKAKVFKDVFLEMNIPYSVVRGEQIQ 840

Db    812 lkatvmnylptsmqmsvqleaspdf-ta-vpvgddqdsyclsangrhtsswlvtpkslgn 869
          ||:|| ||  ::||  | : |    : |:  || | |  :    |    |:          :: |
Qy    841 LKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTKS-SKCVRQKVE--GSSSHL-- 895

Db    870 vnfsvsaeaqqssepcgsevatvpatgrkdtvvkvlivepegikqe-htfsslfcasd-a 927
          | |:| :  : :          : |     | |: :|| | | |||:|:|  :  :|    :  :
Qy    896 VTFTVLP-LEIGLHNINFSLET--WFG-KEILVKTLRVVFEGVKRESYSGVTLDPRGIYG 951

Db    928 eisekmssg-ppp-tvv-kdsarahfsv---m-gdilssai-rnt qnpl-hmpygcgeqn 978
           || ::       |  :| |    :  :||    : |:|||: : ::  | | |:| | :| :
Qy    952 TISRRKEFPYRIPLDLVPKTEIKRILSVKGLLVGEILSAVLSQEGINILTHLPKGSAEAE 1011

Db    979 mvlfapniyvlkyln-etq-ql--tqkiktka-lgf-lragyqrelnykhkdgsysafgd 1032
           ::    | :||: ||:    :   :   :: :   |   |     |: |       ::|:: | |||
Qy   1012 LMSVVPVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEGMLSIMSYRNADYSYSVW-- 1069

Db   1033 qngeregntwltafvlksfaqarafifideshithaftwlsqk-qkdngcfrssgslfnn 1091
           :|   ::||||||  |: ::|    ::   :::  | :::  || ::  |  |||  |: :. | :
Qy   1070 -KGGS-ASTWLTAFALRVLGQVNKYVEQNQNSICNSLLWLVENYQLDNGSFKEN-SQYQP 1126

Db   1092 a-mkggvd-demtlsayit-mallesslp-athpvvskalsclesswktieqernasfvy 1147
            : | :  :     | |:|  :::     :      |:|      : : :    :|     |   :
Qy   1127 IKLQGTLPVEARENSLYLTAFTVIGIRKAFDICPLVKIDTALIKADNFLLENTLPAQSTF 1186

Db   1148 tkalmayafalagnqnkr-deilksldeea-ikennsihwkrpqksrksehhlykpqass 1205
           | |: |||::|: : : :     |: :|   || :|  | :|  : : :: :   :   : | :::
Qy   1187 TLAISAYALSLGDKTHPQFRSIVSALKREALVKGNPPIYRFWKDNLQHKDSSV--PNTGT 1244

Db   1206 ae-vemnayvvlarltaqpapspedltlsmstimwltkqqnsnggfsstqdtvvaldals 1264
           |  ||  ||   |   ||:      :   |:       : |:||: :|      ||| ||||: |:::|:
Qy   1245 ARMVETTAYA-L--LTSL---NLKDINYVNPVIKWLSEEQRYGGGFYSTQDTINAIEGLT 1298

Db   1265 kyga-vtfsrsqkttlvtiqstgsfsqkfqvensnrlllqqvalpdipgdytisvs-qeg 1322
            |:  |    |  :     |:     |:: ::: : : | :| : | :  :   |  :| : | |
Qy   1299 EYSLLVKQLRLSMDIDVSYKHKGAL-HNYKMTDKN-FLGRPVEVL-LNDDLIVSTGFGSG 1355

Db   1323 cvyaq-tmlrynmhlekqlsafaiwvqtvpltcnnpkgh-ns-fqisle-isytgsr--p 1376
               : | :  :       :: :| | ::|  :   :: :|   || :    :    || ||  :
Qy   1356 LATVHVTTVVHKTSTSEEVCSFYLKIDTQDIEASHYRGYGNSDYKRIVACASYKPSREES 1415

Db   1377 as-nm-viadvkmlsgfiplkptvkkl-erlehv-srtevsnnnvliyldqv-tnqtlaf 1431
          :| :    : |: : :|: :     :| | | :::: :   :: : :|:: |: : ::: |
Qy   1416 SSGSSHAVMDISLPTGISANEEDLKALVEGVDQLFTDYQIKDGHVILQLNSIPSSDFLCV 1475
```



```
Db   1432 sfiiqqdipvrnlqpaivkvydyyetd 1458
               | | : : |   | ||    ||:|    |
Qy   1476 RFRIFELFEVGFLSPATFTVYEYHRPD 1502

Search completed: Wed Jan 28 12:17:15 1998
Job time : 69 secs.
```

```
******************************************************************************

MPSRCH (TM)

******************************************************************************
```

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Jan 28 12:12:44 1998; MasPar time 34.21 Seconds
602.393 Million cell updates/sec

Tabular output not generated.

Title: >US-08-487-283A-2
Description: (1-1676) from US08487283A.pep
Perfect Score: 12048
Sequence: 1 MGLLGILCFLIFLGKTWGQE..........CQAFLANLDEFAEDIFLNGC 1676

Scoring table: PAM 150
Gap 11

Searched: 101610 seqs, 12294212 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq28
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21

Statistics: Mean 40.813; Variance 192.520; scale 0.212

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	12048	100.0	1676	15	R77604	Pro-C5 polypeptide.	0.00e+00
2	2369	19.7	1651	12	R63222	Pre-pro-cobra C3 prot	1.94e-197
3	2339	19.4	1663	16	R94030	Human modified C3 (D7	1.02e-194
4	2339	19.4	1663	16	R94028	Human C3 precursor.	1.02e-194
5	2335	19.4	1663	16	R94029	Human modified C3 (R1	2.35e-194
6	2213	18.4	1642	12	R63223	Cobra CVF1 .	2.66e-183
7	1692	14.0	1333	12	R63224	Cobra partial CVF2.	3.18e-136
8	611	5.1	1484	2	R11749	Human alpha-2 macrogl	4.56e-40
9	564	4.7	1474	2	R11334	Recombinant human alp	5.28e-36
10	541	4.5	74	2	P71666	Human anaphylatoxin C	5.04e-34
11	541	4.5	74	13	R75497	Human C5a protein.	5.04e-34
12	540	4.5	403	13	R71933	Cobra venom factor C-	6.15e-34
13	450	3.7	344	4	R21775	Phospholipase A2 inhi	2.98e-26
14	436	3.6	344	3	R10899	Rat phospholipase A2	4.57e-25
15	429	3.6	344	10	R51948	Phospholipase A2 inhi	1.78e-24
16	438	3.6	349	10	R51949	Phospholipase A2 inhi	3.10e-25
17	438	3.6	349	4	R21776	Phospholipase A2 inhi	3.10e-25
18	438	3.6	349	3	R10900	Human phospholipase A	3.10e-25
19	141	1.2	21	15	R77605	Pro-C5 polypeptide KS	1.42e-01
20	139	1.2	24	2	P71532	Sequence of new pepti	1.96e-01
21	116	1.0	23	2	P71534	Sequence of new pepti	6.86e+00
22	105	0.9	24	2	P71533	Sequence of new pepti	3.43e+01
23	105	0.9	153	11	R60520	Human alpha-2-MRBDv.	3.43e+01
24	109	0.9	648	16	R94906	RAP-1 radiation prote	1.93e+01
25	110	0.9	803	12	R60882	P. abyssi DNA polymer	1.67e+01
26	104	0.9	803	12	R60883	P. occultum DNA polym	3.96e+01
27	104	0.9	803	16	R92523	Pyrodictium occultum	3.96e+01
28	110	0.9	903	16	R87007	Hyperthermostable pro	1.67e+01
29	110	0.9	1398	16	R87008	Protease.	1.67e+01
30	100	0.8	361	2	R11547	T. hyo 39 kD family 2	6.98e+01
31	102	0.8	394	21	W20797	H. pylori transporter	5.27e+01
32	100	0.8	465	17	R88328	Cotton UDP glucose sy	6.98e+01
33	96	0.8	522	19	W06330	Receptor-type tyrosin	1.22e+02
34	101	0.8	581	14	R74187	Chick p75.	6.06e+01
35	101	0.8	601	15	R90696	Borrelia burgdorferi	6.06e+01
36	101	0.8	618	15	R90699	Borrelia burgdorferi	6.06e+01
37	100	0.8	628	12	R64979	Pig Syk.	6.98e+01
38	101	0.8	811	14	R72737	Plasmodium falciparum	6.06e+01
39	96	0.8	972	19	W06331	Full-length receptor-	1.22e+02
40	96	0.8	987	14	R85930	Protein tyrosine-kina	1.22e+02
41	96	0.8	987	19	W06335	Full length receptor-	1.22e+02
42	96	0.8	987	20	W11304	Receptor-type tyrosin	1.22e+02
43	98	0.8	1003	5	R29648	AmEPV Spheroidin prot	9.22e+01
44	98	0.8	1003	10	R55576	AmEPV spheroidin.	9.22e+01
45	98	0.8	2199	16	R94562	Human cytotactin.	9.22e+01

ALIGNMENTS

```
RESULT   1
ID   R77604 standard; Protein; 1676 AA.
AC   R77604;
DT   15-MAR-1996  (first entry)
DE   Pro-C5 polypeptide.
KW   Complement C5; haemolysis; kidney; glomerulonephritis;
KW   monoclonal antibody; antiinflammatory; antibody engineering;
KW   humanised antibody.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Peptide         1..18
FT   /label= Sig_peptide
FT   Protein         19..673
FT   /label= Beta-chain
FT   Cleavage_site   673..674
FT   Cleavage_site   677..678
FT   Peptide         674..677
FT   label= Cleavage_peptide
FT   Protein         678..1676
FT   /label= Alpha-chain
FT   /note= "amino acids 872-892 (854-874 of
FT   the mature protein) comprise the KSSKS
FT   epitope"
FT   Peptide         678..751
FT   /label= C5a
FT   Cleavage_site   751..752
```

```
FT   /label= Convertase_cleavage_site
FT   Modified_site   911
FT   /label= N-glycosylation_site
FT   Modified_site   1115
FT   /label= N-glycosylation_site
FT   Modified_site   1630
FT   /label= N-glycosylation_site
PN   WO9529697-A1.
PD   09-NOV-1995.
PF   01-MAY-1995; U05688.
PR   02-MAY-1994; US-236208.
PA   (ALEX-) ALEXION PHARM INC.
PI   Evans MJ,  Matis L,  Mueller EE,  Nye SH,  Rollins S;
PI   Rother RP,  Springhorn J P,  Squinto SP,  Thomas TC;
PI   Wang Y,  Wilkins JA;
DR   WPI; 95-392923/50.
PT   Treating glomerulonephritis with antibody against complement C5
PT   component  - to inhibit complement induced cell lysis
PS   Example 13; Page 82-92; 181pp; English.
CC   The cDNA sequence of the complement C5 gene transcript predicts a
CC   secreted pro-C5 precursor of 1676 amino acids (R77604).  C5 is a
CC   beta-globulin heterodimer thought to play a role in the pathogenesis
CC   of glomerulonephritis (GN).  Cleavage of the C5 alpha-chain
CC   by a convertase enzyme generates anaphylatoxic C5a.  Monoclonal
CC   and humanised recombinant antibodies that recognise the alpha-chain
CC   KSSKC epitope (R77605) block C5a generation, thereby reducing
CC   glomerular inflammation and kidney dysfunction associated with GN.
SQ   Sequence   1676 AA;

  Query Match            100.0%;  Score 12048;  DB 15;  Length 1676;
  Best Local Similarity 100.0%;  Pred. No. 0.00e+00;
  Matches  1676;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Db      1 mgllgilcfliflgktwgqeqtyvisapkifrvgasenivivqvygyteafdatisiksyp 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 MGLLGILCFLIFLGKTWGQEQTYVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSYP 60

Db     61 dkkfsyssghvhlssenkfqnsailtiqpkqlpggqnpvsyvylevvskhfskskrmpit 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 DKKFSYSSGHVHLSSENKFQNSAILTIQPKQLPGGQNPVSYVYLEVVSKHFSKSKRMPIT 120

Db    121 ydngflfihtdkpvytpdqsvkvrvyslnddlkpakretvltfidpegsevdmveeidhi 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    121 YDNGFLFIHTDKPVYTPDQSVKVRVYSLNDDLKPAKRETVLTFIDPEGSEVDMVEEIDHI 180

Db    181 giisfpdfkipsnprygmwtikakykedfsttgtayfevkeyvlphfsvsiepeynfigy 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    181 GIISFPDFKIPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEPEYNFIGY 240

Db    241 knfknfeitikaryfynkvvteadvyitfgiredlkddqkemmqtamqntmlingiaqvt 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    241 KNFKNFEITIKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLINGIAQVT 300

Db    301 fdsetavkelsyysledlnnkylyiavtviestggfseeaeipgikyvlspyklnlvatp 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    301 FDSETAVKELSYYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLNLVATP 360

Db    361 lflkpgipypikvqvkdsldqlvggvpvilnaqtidvnqetsdldpsksvtrvddgvasf 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    361 LFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDDGVASF 420

Db    421 vlnlpsgvtvlefnvktdapdlpeenqaregyraiayssls qsylyidwtdnhkallvge 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    421 VLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKALLVGE 480
```



```
Db    481 hlniivtpkspyidkithynylilskgkiihfgtrekfsdasyqsinipvtqnmvpssrl 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    481 HLNIIVTPKSPYIDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQNMVPSSRL 540

Db    541 lvyyivtgeqtaelvsdsvwlnieekcgnqlqvhlspdadayspgqtvslnmatgmdswv 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    541 LVYYIVTGEQTAELVSDSVWLNIEEKCGNQLQVHLSPDADAYSPGQTVSLNMATGMDSWV 600

Db    601 alaavdsavygvqrgakkplervfqfleksdlgcgaggglnnanvfhlagltfltnanad 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    601 ALAAVDSAVYGVQRGAKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLAGLTFLTNANAD 660

Db    661 dsqendepckeilrprrtlqkkieeiaakykhsvvkkccydgacvnndetceqraarisl 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    661 DSQENDEPCKEILRPRRTLQKKIEEIAAKYKHSVVKKCCYDGACVNNDETCEQRAARISL 720

Db    721 gprcikafteccvvasqlranishkdmqlgrlhmktllpvskpeirsyfpeswlwevhlv 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    721 GPRCIKAFTECCVVASQLRANISHKDMQLGRLHMKTLLPVSKPEIRSYFPESWLWEVHLV 780

Db    781 prrkqlqfalpdslttweiqgigisntgicvadtvkakvfkdvflemnipysvvrgeqiq 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    781 PRRKQLQFALPDSLTTWEIQGIGISNTGICVADTVKAKVFKDVFLEMNIPYSVVRGEQIQ 840

Db    841 lkgtvynyrtsgmqfcvkmsavegictsespvidhqgtksskcvrqkvegssshlvtftv 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    841 LKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTKSSKCVRQKVEGSSSHLVTFTV 900

Db    901 lpleiglhninfsletwfgkeilvktlrvvpegvkresysgvtldprgiygtisrrkefp 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    901 LPLEIGLHNINFSLETWFGKEILVKTLRVVPEGVKRESYSGVTLDPRGIYGTISRRKEFP 960

Db    961 yripldlvpkteikrilsvkgllvgeilsavlsqeginilthlpkgsaeaelmsvvpvfy 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    961 YRIPLDLVPKTEIKRILSVKGLLVGEILSAVLSQEGINILTHLPKGSAEAELMSVVPVFY 1020

Db   1021 vfhyletgnhwnifhsdpliekqklkkklkegmlsimsyrnadysysvwkggsastwlta 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1021 VFHYLETGNHWNIFHSDPLIEKQKLKKKLKEGMLSIMSYRNADYSYSVWKGGSASTWLTA 1080

Db   1081 falrvlgqvnkyveqnqnsicnsllwlvenyqldngsfkensqyqpiklqgtlpvearen 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1081 FALRVLGQVNKYVEQNQNSICNSLLWLVENYQLDNGSFKENSQYQPIKLQGTLPVEAREN 1140

Db   1141 slyltaftvigirkafdicplvkidtalikadnfllentlpaqstftlaisayalslgdk 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1141 SLYLTAFTVIGIRKAFDICPLVKIDTALIKADNFLLENTLPAQSTFTLAISAYALSLGDK 1200

Db   1201 thpqfrsivsalkrealvkgnppiyrfwkdnlqhkdssvpntgtarmvettayalltsln 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1201 THPQFRSIVSALKREALVKGNPPIYRFWKDNLQHKDSSVPNTGTARMVETTAYALLTSLN 1260

Db   1261 lkdinyvnpvikwlseeqryggfystqdtinaiegltkeysllvkqlrlsmdidvsykhk 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1261 LKDINYVNPVIKWLSEEQRYGGFYSTQDTINAIEGLTEYSLLVKQLRLSMDIDVSYKHK 1320

Db   1321 galhnykmtdknflgrpvevllnddlivstgfgsglatvhvttvvhktstseevcsfylk 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1321 GALHNYKMTDKNFLGRPVEVLLNDDLIVSTGFGSGLATVHVTTVVHKTSTSEEVCSFYLK 1380

Db   1381 idtqdieashyrgygnsdykrivacasykpsreesssgsshavmdislptgisaneedlk 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1381 IDTQDIEASHYRGYGNSDYKRIVACASYKPSREESSSGSSHAVMDISLPTGISANEEDLK 1440
```

```
Db   1441 alvegvdqlftdyqikdghvilqlnsipssdflcvrfrifelfevgflspatftvyeyhr 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1441 ALVEGVDQLFTDYQIKDGHVILQLNSIPSSDFLCVRFRIFELFEVGFLSPATFTVYEYHR 1500

Db   1501 pdkqctmfystsnikiqkvcegaackcveadcgqmqeeldltisaetrkqtackpeiaya 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1501 PDKQCTMFYSTSNIKIQKVCEGAACKCVEADCGQMQEELDLTISAETRKQTACKPEIAYA 1560

Db   1561 ykvsitsitvenvfvkykatlldiyktgeavaekdseitfikkvtctnaelvkgrqylim 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1561 YKVSITSITVENVFVKYKATLLDIYKTGEAVAEKDSEITFIKKVTCTNAELVKGRQYLIM 1620

Db   1621 gkealqikynfsfryiypldsltwieywprdttcsscqaflanldefaedifl ngc 1676
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1621 GKEALQIKYNFSFRYIYPLDSLTWIEYWPRDTTCSSCQAFLANLDEFAEDIFLNGC 1676
```

```
RESULT   2
ID   R63222 standard; Protein; 1651 AA.
AC   R63222;
DT   30-JUN-1995  (first entry)
DE   Pre-pro-cobra C3 protein.
KW   Cobra; C3; third component of complement; human; mouse; rat;
KW   X. laevis; pre-pro molecule; beta chain; alpha chain; codon usage;
KW   G+C content; immune response; host defence; glycosylation;
KW   tertiary structure; maturation; thioester site;
KW   alternative pathway activation.
OS   Naja naja.
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT   /note= "Signal sequence"
FT   Protein         23..655
FT   /note= "Alpha-chain"
FT   Modified_site   153
FT   /note= "Potential glycosylation site"
FT   Modified_site   158
FT   /note= "Potential glycosylation site"
FT   Modified_site   210
FT   /note= "Potential glycosylation site"
FT   Protein         739..1269
FT   /note= "Gamma-chain"
FT   Binding_site    742..751
FT   /note= "Factor B binding site"
FT   Active_site     996..1017
FT   /note= "Thioester site"
FT   Binding_site    1176..1201
FT   /note= "Factor H binding site"
FT   Binding_site    1214..1219
FT   /note= "Potential CR2 binding site"
FT   Binding_site    1221..1235
FT   /note= "Factor H binding site"
FT   Protein         1270..1651
FT   /note= "Beta-chain"
FT   Modified_site   1352
FT   /note= "Potential glycosylation site"
FT   Binding_site    1419..1442
FT   /note= "Properdin binding site"
PN   WO9423024-A.
PD   13-OCT-1994.
PF   07-APR-1994; U03441.
PR   07-APR-1993; US-043747.
PA   (GEOU ) UNIV GEORGETOWN.
PI   Bredehorst R,  Fritzinger DC,  Vogel C;
```



```
DR   WPI; 94-333186/41.
DR   N-PSDB; Q77789.
PT   DNA encoding cobra C3, CVF 1 and CVF 2 - which are used in the
PT   treatment of cancer
PS   Claim 1; Fig 2A-2L; 155pp; English.
CC   This sequence represents cobra C3 (third component of complement).
CC   The cDNA sequence of cobra C3 shows a high sequence homology with C3
CC   molecules from human, mouse, rat and X. laevis.  Cobra C3 is
CC   synthesised as a pre-pro molecule that is subsequently processed
CC   into the mature two-chain protein by removing the signal peptide and
CC   the four Arg residues between the beta and alpha chain.  The alpha
CC   chain comprises 992 amino acids and the beta chain comprises 633
CC   residues, being 12 residues shorter than the human beta chain.  In
CC   contrast to C3 molecules studied so far, cobra C3 is not glycosylated.
CC   All 27 Cys residues are conserved indicating a very high degree of
CC   similarity in the tertiary structure of mammalian and cobra C3.  There
CC   are four Arg residues in pre-pro-C3 which are required for the proper
CC   maturation to yeild the secreted C3 molecule, and the thioester site,
CC   which is unique to the alpha chain, is required for alternative pathway
CC   activation and covalent binding to target cells. Cobra C3 has a
CC   different codon usage compared to mammalian C3 mRNAs.  The G+C content
CC   of all known mammalian C3 mRNAs is more than 53%.  The G+C content of
CC   cobra C3 mRNA is significantly lower at 43%.  The significance of this
CC   difference is not known.  C3 is thought to have important functions in
CC   the immune response and host defence.
SQ   Sequence   1651 AA;

  Query Match             19.7%;  Score 2369;  DB 12;  Length 1651;
  Best Local Similarity   30.2%;  Pred. No. 1.94e-197;
  Matches  518;  Conservative  416;  Mismatches  655;  Indels  127;  Gaps  98;

Db      4 malylvaalligfpg-s-shgalytlitpavlrtdteeqilveahgdstpksldifvhdf 61
          |:|  :  :|| | | : ::   | : :| ::|  : | |:::  | : :     | :  :
Qy      1 MGLLGILCFLI-FLGKTWGQEQTYVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSY 59

Db     62 prkqktlfqsrvdmnqagsmfvtpti-kvpakelnkdskqnqyvvvkvtgpqvalekvvl 120
          | |  :    ::| :: : ::|  ::|    : :|:|    :   || : | : : :  | :
Qy     60 PDKKFSYSSGHVHLS-SENKFQNSAILTIQPKQLPGGQNPVSYVYLEVVSKHFSKSKRMP 118

Db    121 lsyqsgfvfiqtdkgiytpgspvryrvfsvdhnmhrmdktvivefqtpegivvs-skpvn 179
          ::|::||:||:||| :||| :|: ||:|:: ::     :  :: |  |||  |       ::
Qy    119 ITYDNGFLFIHTDKPVYTPDQSVKVRVYSLNDDLKPAKRETVLTFIDPEGSEVDMVEEID 178

Db    180 psgsir-p-ynlpelvsfgtwkavakyehspeesytayfdvreyvlpsfevrlqpsdkfl 237
             | |  | :::|     :| |    |||        : ||||:|:||||| | | ::|  :|:
Qy    179 HIGIISFPDFKIPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEPEYNFI 238

Db    238 -yidgnknfhvsitarylygkkv-egvafvvfgvk--i-ddakksipdsltripiidgdg 292
           | :  ||| ::| |||:| | | |:  :: ||::  : || |   :  ::     :|:| :
Qy    239 GYKN-FKNFEITIKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLINGIA 297

Db    293 eatlkrdt-lrs-rfqdlnqlvghtlyvsvtvitesgsdmvvteqggihivtspyqiyft 350
          : |:  :| ::   :  |::|    ||::||||  :|:     :|  ||  | ||| : :
Qy    298 QVTFDSETAVKELSYYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLNLV 357

Db    351 ktpkyfkpgmpyeltvyvtnp-d---gs-paa-hvp-v-vs-ea--ih-segtt-lsdgt 397
          || ::|||:|| : | | :: |    |: |     :     : |: |:   :   | : | : ||
Qy    358 ATPLFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDDGV 417

Db    398 aklilntplniqslpitvrtnhgdlprerqaiksmtatayqtqggsenylhvaitsteik 457
          | ::|| |   :   | : |:|:  ||| | ||   :  | || : : |   |:  :       :
Qy    418 ASFVLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKALL 477

Db    458 pgdnlpvnfnvrgnanslnqikyftylilnkgkifkvgrqprrdgqnlvtmnlhitpdli 517
```

```
               |::| : : |  ::  :: |  : ||||:||||:  | : :    :  ::|: :| :::
Qy     478 VGEHLNI-I-VTPKSPYIDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQNMV 535

Db     518 psfrfvayyqv-gnn--eivadsvwvdvkdtcmgtlvvkgassrddriqkpgaamkikle 574
               || |:: || | |:    |:|:||||::: : |  | |    |:  |     || :: :::
Qy     536 PSSRLLVYYIVTGEQTAELVSDSVWLNIEEKCGNQLQVH-LSPDADAY-SPGQTVSLNMA 593

Db     575 gdpgarvglvavdkavyvlndkykisqakiwdtieksdfgctagsgqnnlgvfedaglal 634
                 ::|:| ||| ||| :     | :  :: : :||||:|| ||:| ||  ||  |||::
Qy     594 TGMDSWVALAAVDSAVYGVQRGAKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLAGLTF 653

Db     635 ttstnlntkqrsaakcpqpanrrrrssvllldskaskaaqfqdqglrkccedgmhdnpmg 694
               |::| :  | :    | :    | ||:    |          || :     ::||| ||    |
Qy     654 LTNANADDSQENDEPCKE-ILRPRRT---LQKKIEEIAAKYKHSVVKKCCYDGACVNNDE 709

Db     695 ytcekrakyiqegdackaaflecchyikgirdenqreselflarsdfedelfgddniisr 754
               ||| ||   |   |   |   || |||       :|  |      :: |:|         |   | |
Qy     710 -TCEQRAARISLGPRCIKAFTECCVVASQLRA-NISHKDMQLGRLHMKTLLPVSKPEI-R 766

Db     755 sdfpeswlwlteeltgepnnqgisstkvpfylrdsittwellavglsptkgicvaepyei 814
               | |||||||      |      |     :        | : | |  ||:||||: ::|:| | ||||:
Qy     767 SYFPESWLWEVH-LV--PR-R----KQLQFALPDSLTTWEIQGIGISNT-GICVADTVKA 817

Db     815 tvmkdffidlrlpysvvkneqveirailynyadedi-y-vr---veliynpafcsasteg 869
               | || |::: :|||||: ||:::::: :|||    : : |:   || |   :         :|
Qy     818 KVFKDVFLEMNIPYSVVRGEQIQLKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQG 877

Db     870 qry-r-qqfpikalssravpfvivpleqglhdveviasvrgelasdgvrkklkvvpeger 927
               :   :   :  :  : :  |: | | :: ||| |||:::        :   :: : : | |:||||| :
Qy     878 TKSSKCVRQKVEGSSSHLVTFTVLPLEIGLHNINFSLETW--FGKEILVKTLRVVPEGVK 935

Db     928 knivtiieldpsvkgv-ggtqeltviankld-dkvpdtevetrisvlgdpvaqiiensid 985
               ::  : : |||  :|: |          :: ::   | || ||:    :|| |  |::|:    :
Qy     936 RESYSGVTLDP--RGIYGTISRRKEFPYRIPLDLVPKTEIKRILSVKGLLVGEILSAVLS 993

Db     986 gsklnhliitpsgcgeqnmitmtpsviaty-yldatgqwenlgvdrrte-a-ikqimt-g 1041
                 :| |    | | :| ::::: | |:  : ||::  :|: :  |    |     :| :  |
Qy     994 QEGINILTHLPKGSAEAELMSVVP-VFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEG 1052

Db    1042 yaqqmvykkadhsyaaftnrassswltayvvkvlamasnmvkdisheiicggvkwlilnr 1101
                 | |::|| ||:       ::|:||||: ::||:  :: | : ::: || :: ||: |
Qy    1053 MLSIMSYRNADYSYSVWKGGSASTWLTAFALRVLGQVNKYV-EQNQNSICNSLLWLVENY 1111

Db    1102 qqpdgvfkenapvihgemlggtkgaep-easl--tafivtallesrsvckeqinildssi 1158
               |  :| ||||:    :   :| ||    |: | ||  ||| | ::   :   :|    ::| |:::
Qy    1112 QLDNGSFKENSQ-YQPIKLQGTLPVEARENSLYLTAFTVIGIRKAFDIC-PLVKI-DTAL 1168

Db    1159 nkatdyllkkyeklqrpyttaltayalaaadrlndd-r--v--lm-aa-stgrn---r-w 1207
               || ::|| :      |  :| |::||||: :|: : : :  |   |  |:   |    |       | |
Qy    1169 IKADNFLLENTLPAQSTFTLAISAYALSLGDKTHPQFRSIVSALKREALVKGNPPIYRFW 1228

Db    1208 -e--ey-narthn------iegtsyallallkmkkfaevgpvvrwlidqkyyggtygqtq 1257
                 :   :  ::    |        :| |:||||: |::| :   | ||::|| ::   ||| :  ||
Qy    1229 KDNLQHKDSSVPNTGTARMVETTAYALLTSLNLKDINYVNPVIKWLSEEQRYGGGFYSTQ 1288

Db    1258 atvmvfqalaeyeiqmpthqdlnldisiklperevperysindrnavqartvetklnedf 1317
               |:  :::|:|| : :  :  |::|| :      :          | : |:| : :| ||   ||:|:
Qy    1289 DTINAIEGLTEYSL-LVKQLRLSMDIDVSYKHKGALHNYKMTDKNFL-GRPVEVLLNDDL 1346

Db    1318 tvsas-gdgkatmtiltvynaqlredanvcnkfhldvsvenvelnlkqakggkaalrlki 1376
               ||:: | | ||: : || :         :||: | | :  :::| :  :: |       |:
Qy    1347 IVSTGFGSGLATVHVTTVVH-KTSTSEEVCS-FYLKIDTQDIEASHYRGYGNSDYKRIVA 1404

Db    1377 ctry-lg--evds-tm-tiidismltgffpdaedlkrlsngvdryiskfeidnnmaqkgt 1431
```



```
               |: |  :  |  | :  :::|||: ||: :: |||| | :|||: :: ::| :: ::
Qy    1405 CASYKPSREESSSGSSHAVMDISLPTGISANEEDLKALVEGVDQLFTDYQI-KD-GH--- 1459

Db    1432 vviyldkvshsedeclhfkihkhfevgfiqpgsvkvysyynldeqctkfyhpdketgvln 1491
               |:: |: :: |:  |::|:|   |||||: |::  || |    | |||:||    :    :
Qy    1460 VILQLNSIPSSDFLCVRFRIFELFEVGFLSPATFTVYEYHRPDKQCTMFYSTSNIK-I-Q 1517

Db    1492 kichgnicrcaeetcsllnqq-k-kidlqlriqkacaqnvdyvyktkllrieekdgndiy 1549
               |:| |  |:| |  |: : ::     |  : | | ||  :: | ||  :  |    :    |
Qy    1518 KVCEGAACKCVEADCGQMQEELDLTISAETRKQTACKPEIAYAYKVSITSITVENVFVKY 1577

Db    1550 fmdvlevikggtdrnaqakarqyvsqrkcqealnlkldndyliwglssdlwpmk-d-dis 1607
                  :|:: | |               ::      | :| :|     :||| |     | :| :  :
Qy    1578 KATLLDIYKTGEAVAEKDSEITFIKKVTCTNA-ELVKGRQYLIMG-KEAL-QIKYNFSFR 1634

Db    1608 ylit-kn-twierwpnedecqd-eef-qnlcddfaq 1639
               |:    : |||| || :  |   :  |  || |:||:
Qy    1635 YIYPLDSLTWIEYWPRDTTCSSCQAFLANL-DEFAE 1669
```

RESULT 3
ID R94030 standard; Protein; 1663 AA.
AC R94030;
DT 21-MAY-1996 (first entry)
DE Human modified C3 (D752G, E753S, D754G).
KW C3 protein; convertase; Factor I; Factor H; complement; tumour;
KW infection; therapy.
OS Synthetic.
FH Key Location/Qualifiers
FT Peptide 1..22
FT /label= Sig_peptide
FT Protein 23..667
FT /note= "C3 beta chain"
FT Peptide 668..671
FT /note= "amino acids 668-671 are removed when the
FT precursor is cleaved into the alpha and
FT beta chains"
FT Protein 672..1663
FT /note= "C3 alpha chain"
PN WO9607738-A2.
PD 14-MAR-1996.
PF 08-SEP-1995; G02121.
PR 08-SEP-1994; GB-018147.
PR 04-MAY-1995; GB-009102.
PA (IMUT-) IMUTRAN LTD.
PI Farries TC, Harrison RA;
DR WPI; 96-171613/17.
PT Mutant complement pathway protein forming stable C3 convertase -
PT for generalised complement depletion or localised complement
PT activation
PS Claim 11; Fig 1; 81pp; English.
CC A modified human C3 protein (R94030) differs from the wild-type
CC (R94028) by substitution of Asp-Glu-Asp at positions 752-754 by
CC Gly-Ser-Gly. It is obtained by site-directed mutagenesis of
CC C3-encoding cDNA (T17738). The modification reduces the
CC interaction of C3b/C3i with Factor H in comparison to wild-type
CC C3. This allows the modified C3 to be used therapeutically to
CC super-active the complement system or the increase a target's
CC (e.g. tumour, pathogen or virus-infected cell) sensitivity to
CC complement-mediated destruction.
SQ Sequence 1663 AA;

Query Match 19.4%; Score 2339; DB 16; Length 1663;
Best Local Similarity 28.3%; Pred. No. 1.02e-194;

```
Matches  488; Conservative  459; Mismatches 660; Indels 115; Gaps  90;

Db      8 sll-ll-llthlplalg-spmysiitpnilrleseetmvleahdaqgdvpvtvtvhdfpg 64
          :|| :| :|  |  :  |    | | :|:|:|: : | :|::         |::: :|
Qy      2 GLLGILCFLIFLGKTWGQEQTYVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSYPD 61

Db     65 kklvlssektvltpatnhmgnvtftipanrefksekgrnkfvtvqatfgtqvve-kvvlv 123
          ||:  ||    |::  :    :  :||  : ::: :       :| ::    : :    | : :
Qy     62 KKFSYSSGHVHLSSENKFQNSAILTIQP-KQLPGGQNPVSYVYLEVV-SKHFSKSKRMPI 119

Db    124 slqsgylfiqtdktiytpgstvlyriftvnhkllpvgrtvmvnienpegipvkqdslssq 183
          : ::|:|||:||| :|||  :|   |::::|  | |  :  :: : :||   |  | :
Qy    120 TYDNGFLFIHTDKPVYTPDQSVKVRVYSLNDDLKPAKRETVLTFIDPEGSEV--DMVEEI 177

Db    184 nqlqvlpls-wdipelvnmgqwkirayyenspqqvfstefevkeyvlpsfeviveptekf 242
          :::|:::::    ||       | | |:| |  :         ::  ||||||||| | | :|| | :
Qy    178 DHIGIISFPDFKIPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEP-E-Y 235

Db    243 yyi-yne-kglevtitarflygkkv-egtafvifgi-qd--geqrislpeslkripiedg 296
          :| |:: | :|:|| ||::| | | |:   :: ||| :|     :|:   :    ::      : :|
Qy    236 NFIGYKNFKNFEITIKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLING 295

Db    297 sgevvlsrkvlldgvqnpraedlvgkslyvsatvilhsgsdmvqaersgipivtspyqih 356
          ::| :     :   :        |||   |  |||:: |||  :|:     :|| :||   |  |||  ::
Qy    296 IAQVTFDSETAVKELSYYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLN 355

Db    357 ftktpkyfkpgmpfdlmvfvtnp-d---gs-pay-rvp-vav-q--ge-dtvqsltq-gd 404
          :  || ::||:|: ::| | :: |    :| |             : | |    :: |     |:|:  |
Qy    356 LVATPLFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDD 415

Db    405 gvaklsinthpsqkplsitvrtkkqelseaeqatrtmqalpystvgnsnnylhlsvlrte 464
          ||| : :|  ::      | : |:|     :|:| :||       :|::||::: |   |:  :  :
Qy    416 GVASFVLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKA 475

Db    465 lrpgetlnvnfllrmdraheakiryytylimnkgrllkagrqvrepgqdlvvlplsittd 524
          |  || ||: ::   :    :    ||   |  |||::|||::   |  : :  :     :  :::| :
Qy    476 LLVGEHLNI-IVTPKS-PYIDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQN 533

Db    525 fipsfrlvayytligasgqrevvadsvwvdvkdscvgslvvksgqsedrqpvpgqqmtlk 584
          :|| ||: || : |    |:|:||||::: : |    |  |    :     |  : ||| ::|:
Qy    534 MVPSSRLLVYYIVTGEQTA-ELVSDSVWLNIEEKCGNQLQVHLSPDADAYS-PGQTVSLN 591

Db    585 iegdhgarvvlvavdkgvfvlnkknkltqskiwdvvekadigctpgsgkdyagvfsdagl 644
          :       ::| | ||| :|: : :   |      :: : :||:|:|| :|:| : | ||   |||
Qy    592 MATGMDSWVALAAVDSAVYGVQRGAKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLAGL 651

Db    645 tftsssgqqtaqraelqcpqpaarrrrsvqltekrmdkvg-kypkel-rkccedgmrenp 702
          || :::  : :|  :   | :    | ||:   | :|::: :: ||    : :||| ||   |
Qy    652 TFLTNANADDSQENDEPCKE-ILRPRRT--L-QKKIEEIAAKYKHSVVKKCCYDGACVNN 707

Db    703 mrfscqrrtrfislgeackkvfldccnyitelrrqharashlglarsnlgsgiiaeeniv 762
          :|::|:  ||||  | | |  :||    ::||    ::   : |:| :: : ::: :
Qy    708 DE-TCEQRAARISLGPRCIKAFTECCVVASQLRANISH-KDMQLGRLHMKT-LLPVSKPE 764

Db    763 srsefpeswlwnvedlkeppkngistklmniflkdsittweilavsmsdkkgicvadpfe 822
          || ||||||||:|   |   | ::       |  :  | ||:|||||  ::::|:  ||||||
Qy    765 IRSYFPESWLWEVH-L-VPRRK----QL-QFALPDSLTTWEIQGIGISNT-GICVADTVK 816

Db    823 vtvmqdffidlrlpssvvrneqveiravlynyrqnqelkvrvellhnpafcslattkrrh 882
           |  | |::: :| ||||  ||::::: :||||  :    :   | :    ::|:   :    |
Qy    817 AKVFKDVFLEMNIPYSVVRGEQIQLKGTVYNYRTS-GMQFCVKMSAVEGICTSESPVIDH 875

Db    883 qqtit---ippk--ssls--vpyvivplktglqevevkaavyhhfisdgvrkslkvvpeg 935
          | | :   :  |  :| |  | : ::||  ||::::          |  : : |:|:|||||
Qy    876 QGTKSSKCVRQKVEGSSSHLVTFTVLPLEIGLHNINFSLETW--FGKEILVKTLRVVPEG 933

Db    936 irmnktvavrtldperl-gregvqkedip-padlsdqvpdtesetril-lqgtpvaqmte 992
          :: :    :| ||||  : |   : :|| :|    | | || ||    ||| : |  |:::
Qy    934 VKRESYSGV-TLDPRGIYGTISRRKE-FPYRIPL-DLVPKTEI-KRILSVKGLLVGEILS 989

Db    993 davdaerlkhlivtpsgcgeqnmigmtptviavhyldeteqwekf-g--lekrqgaleli 1049
           :  | :: |   | | :| ::::: |      |||:  ::|: | :  |  :|      :
Qy    990 AVLSQEGINILTHLPKGSAEAELMSVVPVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKL 1049

Db   1050 kkgytqqlafrqpssafaafvkrapstwltayvvkvfslavnliaidsqvlcgavkwlil 1109
          | |     :::| :  :::       ::||||||: ::|::    : :  :     :| :: ||:
Qy   1050 KEGMLSIMSYRNADYSYSVWKGGSASTWLTAFALRVLGQVNKYVEQNQNSICNSLLWLVE 1109

Db   1110 ekqkpdgvfqedapvihqemigglr-nnnekdmaltafvlislqeakdiceeqvnslpgs 1168
          : |  :| | |::        : | |   :   |: : |||| :|::: | |||   |: :  :
Qy   1110 NYQLDNGSFKENSQYQPIKLQGTLPVEARENSLYLTAFTVIGIRKAFDICP-LVK-IDTA 1167

Db   1169 itkagdfleanymnlqrsytvaiagyala--qmgr--lkg--pllnk-flttakdk-nr- 1219
          : || :||  | :   | ::|:|||::|||:  :: :   :::   : |::   |  ::      |
Qy   1168 LIKADNFLLENTLPAQSTFTLAISAYALSLGDKTHPQFRSIVSALKREALVKGNPPIYRF 1227

Db   1220 w-ed-pgkq--l-yn-----veatsyallallqlkdfdfvppvvrwlneqryygggygst 1269
          | ::   |:  :          ||:|:||||: | |||:::| ||::||:|:: ||||: ||
Qy   1228 WKDNLQHKDSSVPNTGTARMVETTAYALLTSLNLKDINYVNPVIKWLSEEQRYGGGFYST 1287

Db   1270 qatfmvfqalaqyqkdapdhqelnldvslqlpsrsskithrihwesasllrseetkeneg 1329
          | |:  :::|::|         |::|:|: :    :::         ::  :   | |:: |   |:
Qy   1288 QDTINAIEGLTEYSL-LVKQLRLSMDIDVSYKHKGALHNYKMT-DKNFLGRPVEVLLNDD 1345

Db   1330 ftvtae-gkgqgtlsvvtmyhakakdqltcnkfdlkvtikpapetekrpqdakntmilei 1388
          : |::  | | :|:  | |:  |   :   :  : |:  | ||:             :   |     : : :
Qy   1346 LIVSTGFGSGLATVHVTTVVHKTSTSEEVCS-FYLKIDTQDIEASHYRGYGNSDYKRIVA 1404

Db   1389 ccryr-g-dqdat-ms--ildismmtgfapdtddlkqlangvdryiskyeldkafsdrnt 1443
          |: |: : :: ::  |   ::||||: ||:::: :||| | :|||: :: ||:: |  :  :
Qy   1405 CASYKPSREESSSGSSHAVMDISLPTGISANEEDLKALVEGVDQLFTDYQI-KD-G--H- 1459

Db   1444 liiyldkvshseddclafkvhqyfnveliqpgavkvyayynleesctrfyhpekedgkln 1503
          :|: |: :: |:  |: |:: : |:| :: |::  || |   :  || || :   : |:
Qy   1460 VILQLNSIPSSDFLCVRFRIFELFEVGFLSPATFTVYEYHRPDKQCTMFY-S-TSNIKIQ 1517

Db   1504 klcrdelcrcaeencf-iqksdd-kvtleerldkacepgvdyvykt rlvkvqlsndfdey 1561
          |:|    |:| | :|  :|    |   :: | | : || | : | || :  : : | |  |
Qy   1518 KVCEGAACKCVEADCGQMQEELDLTISAETRKQTACKPEIAYAYKVSITSITVENVFVKY 1577

Db   1562 imaieqtiksgsdevqvgqqrtfispikcrealkleekkhylmwglss-df-wgekpnls 1619
            ::: |:|    :  :|||  : | :|  |    ::||: |  : ::
Qy   1578 KATLLDIYKTGEAVAEKDSEITFIKKVTCTNA-ELVKGRQYLIMGKEALQIKYNFSFRYI 1636

Db   1620 yiigkdtwvehwpeedecqdeenqkqcqdlgaftesmvvfgc 1661
          | :   ||:| || :   |    :       :|  |:| : : ||
Qy   1637 YPLDSLTWIEYWPRDTTCSSCQAF-L-ANLDEFAEDIFLNGC 1676
```

```
RESULT    4
ID   R94028 standard; Protein; 1663 AA.
AC   R94028;
DT   21-MAY-1996  (first entry)
DE   Human C3 precursor.
KW   C3 protein; convertase; Factor I; Factor H; complement.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Peptide         1..22
```

```
FT   /label= Sig_peptide
FT   Protein         23..667
FT   /note= "C3 beta chain"
FT   Peptide         668..671
FT   /note= "amino acids 668-671 are removed when the
FT   precursor is cleaved into the alpha and
FT   beta chains"
FT   Protein         672..1663
FT   /note= "C3 alpha chain"
PN   WO9607738-A2.
PD   14-MAR-1996.
PF   08-SEP-1995; G02121.
PR   08-SEP-1994; GB-018147.
PR   04-MAY-1995; GB-009102.
PA   (IMUT-) IMUTRAN LTD.
PI   Farries TC,  Harrison RA;
DR   WPI; 96-171613/17.
DR   N-PSDB; T17738.
PT   Mutant complement pathway protein forming stable C3 convertase  -
PT   for generalised complement depletion or localised complement
PT   activation
PS   Disclosure; Fig 1; 81pp; English.
CC   Human C3 protein (R94028) was produced by expression of a cDNA
CC   sequence (T17738) isolated from a human liver cDNA library.
CC   C3 is a complement pathway protein that is succeptible to cleavage
CC   by Factor I and is also susceptible to the inhibitory action
CC   of Factor H.  Mutants of C3 (R94029 and R94030) have been
CC   produced by site-directed mutagenesis.  These mutants can be
CC   used to super-active the complement system, or to induce
CC   localised super-activation at a specific target to increase
CC   the target's sensitivity to complement-mediated destruction.
SQ   Sequence   1663 AA;

  Query Match            19.4%;  Score 2339;  DB 16;  Length 1663;
  Best Local Similarity  28.3%;  Pred. No. 1.02e-194;
  Matches  488;  Conservative  458;  Mismatches 661;  Indels 115;  Gaps  90;

Db       8 sll-ll-llthlplalg-spmysiitpnilrleseetmvleahdaqgdvpvtvtvhdfpg 64

Qy       2 GLLGILCFLIFLGKTWGQEQTYVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSYPD 61

Db      65 kklvlssektvltpatnhmgnvtftipanrefksekgrnkfvtvqatfgtqvve-kvvlv 123

Qy      62 KKFSYSSGHVHLSSENKFQNSAILTIQP-KQLPGGQNPVSYVYLEVV-SKHFSKSKRMPI 119

Db     124 slqsgylfiqtdktiytpgstvlyriftvnhkllpvgrtvmvnienpegipvkqdslssq 183

Qy     120 TYDNGFLFIHTDKPVYTPDQSVKVRVYSLNDDLKPAKRETVLTFIDPEGSEV--DMVEEI 177

Db     184 nqlgvlpls-wdipelvnmgqwkirayyenspqqvfstefevkeyvlpsfeviveptekf 242

Qy     178 DHIGIISFPDFKIPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEP-E-Y 235

Db     243 yyi-yne-kglevtitarflygkkv-egtafvifgi-qd--geqrislpeslkripiedg 296

Qy     236 NFIGYKNFKNFEITIKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLING 295

Db     297 sgevvlsrkvlldgvqnpraedlvgkslyvsatvilhsgsdmvqaersgipivtspyqih 356

Qy     296 IAQVTFDSETAVKELSYYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLN 355

Db     357 ftktpkyfkpgmpfdlmvfvtnp-d---gs-pay-rvp-vav-q--ge-dtvqsltq-gd 404

Qy     356 LVATPLFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDD 415
```



```
Db     405 gvaklsinthpsqkplsitvrtkkqelseaeqatrtmqalpystvgnsnnylhlsvlrte 464

Qy     416 GVASFVLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKA 475

Db     465 lrpgetlnvnfllrmdraheakiryytylimnkgrllkagrqvrepgqdlvvlplsittd 524

Qy     476 LLVGEHLNI-IVTPKS-PYIDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQN 533

Db     525 fipsfrlvayytligasgqrevvadsvwvdvkdscvgslvvksgqsedrqpvpgqqmtlk 584

Qy     534 MVPSSRLLVYYIVTGEQTA-ELVSDSVWLNIEEKCGNQLQVHLSPDADAYS-PGQTVSLN 591

Db     585 iegdhgarvvlvavdkgvfvlnkknkltqskiwdvvekadigctpgsgkdyagvfsdagl 644

Qy     592 MATGMDSWVALAAVDSAVYGVQRGAKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLAGL 651

Db     645 tftsssgqqtaqraelqcpqpaarrrrsvqltekrmdkvg-kypkel-rkccedgmrenp 702

Qy     652 TFLTNANADDSQENDEPCKE-ILRPRRT--L-QKKIEEIAAKYKHSVVKKCCYDGACVNN 707

Db     703 mrfscqrrtrfislgeackkvfldccnyitelrrqharashlglarsnldediiaeeniv 762

Qy     708 DE-TCEQRAARISLGPRCIKAFTECCVVASQLRANISH-KDMQLGRLHMK-TLLPVSKPE 764

Db     763 srsefpeswlwnvedlkeppkngistklmniflkdsittweilavsmsdkkgicvadpfe 822

Qy     765 IRSYFPESWLWEVH-L-VPRRK----QL-QFALPDSLTTWEIQGIGISNT-GICVADTVK 816

Db     823 vtvmqdffidlrlpssvvrneqveiravlynyrqnqelkvrvellhnpafcslattkrrh 882

Qy     817 AKVFKDVFLEMNIPYSVVRGEQIQLKGTVYNYRTS-GMQFCVKMSAVEGICTSESPVIDH 875

Db     883 qqtit---ippk--ssls--vpyvivplktglqevevkaavyhhfisdgvrkslkvvpeg 935

Qy     876 QGTKSSKCVRQKVEGSSSHLVTFTVLPLEIGLHNINFSLETW--FGKEILVKTLRVVPEG 933

Db     936 irmnktvavrtldperl-gregvqkedip-padlsdqvpdtesetril-lqgtpvaqmte 992

Qy     934 VKRESYSGV-TLDPRGIYGTISRRKE-FPYRIPL-DLVPKTEI-KRILSVKGLLVGEILS 989

Db     993 davdaerlkhlivtpsgcgeqnmigmtptviavhyldeteqwekf-g--lekrqgaleli 1049

Qy     990 AVLSQEGINILTHLPKGSAEAELMSVVPVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKL 1049

Db    1050 kkgytqqlafrqpssafaafvkrapstwltayvvkvfslavnliaidsqvlcgavkwlil 1109

Qy    1050 KEGMLSIMSYRNADYSYSVWKGGSASTWLTAFALRVLGQVNKYVEQNQNSICNSLLWLVE 1109

Db    1110 ekqkpdgvfqedapvihqemigglr-nnnekdmaltafvlislqeakdiceeqvnslpgs 1168

Qy    1110 NYQLDNGSFKENSQYQPIKLQGTLPVEARENSLYLTAFTVIGIRKAFDICP-LVK-IDTA 1167

Db    1169 itkagdflsanymnlqrsytvaiagyala--qmgr--lkg--pllnk-flttakdk-nr- 1219

Qy    1168 LIKADNFLLENTLPAQSTFTLAISAYALSLGDKTHPQFRSIVSALKREALVKGNPPIYRF 1227

Db    1220 w-ed-pgkq--l-yn-----veatsyallallqlkdfdfvppvvrwlneqryyggygst 1269

Qy    1228 WKDNLQHKDSSVPNTGTARMVETTAYALLTSLNLKDINYVNPVIKWLSEEQRYGGGFYST 1287

Db    1270 qatfmvfqalaqyqkdapdhqelnldvslqlpsrsskithrihwesasllrseetkeneg 1329

Qy    1288 QDTINAIEGLTEYSL-LVKQLRLSMDIDVSYKHKGALHNYKMT-DKNFLGRPVEVLLNDD 1345
```

```
Db    1330 ftvtae-gkgqgtlsvvtmyhakakdqltcnkfdlkvtikpapetekrpqdakntmilei 1388

Qy    1346 LIVSTGFGSGLATVHVTTVVHKTSTSEEVCS-FYLKIDTQDIEASHYRGYGNSDYKRIVA 1404

Db    1389 ctryr-g-dqdat-ms--ildismmtgfapdtddlkqlangvdryiskyeldkafsdrnt 1443

Qy    1405 CASYKPSREESSSGSSHAVMDISLPTGISANEEDLKALVEGVDQLFTDYQI-KD-G--H- 1459

Db    1444 liiyldkvshseddclafkvhqyfnveliqpgavkvyayynleesctrfyhpekedgkln 1503

Qy    1460 VILQLNSIPSSDFLCVRFRIFELFEVGFLSPATFTVYEYHRPDKQCTMFY-S-TSNIKIQ 1517

Db    1504 klcrdelcrcaeencf-iqksdd-kvtleerldkacepgvdyvyktrlvkvqlsndfdey 1561

Qy    1518 KVCEGAACKCVEADCGQMQEELDLTISAETRKQTACKPEIAYAYKVSITSITVENVFVKY 1577

Db    1562 imaieqtiksgsdevqvgqqrtfispikcrealkleekkhylmwglss-df-wgekpnls 1619

Qy    1578 KATLLDIYKTGEAVAEKDSEITFIKKVTCTNA-ELVKGRQYLIMGKEALQIKYNFSFRYI 1636

Db    1620 yiigkdtwvehwpeedecqdeenqkqcqdlgaftesmvvfgc 1661

Qy    1637 YPLDSLTWIEYWPRDTTCSSCQAF-L-ANLDEFAEDIFLNGC 1676
```

```
RESULT   5
ID   R94029 standard; Protein; 1663 AA.
AC   R94029;
DT   21-MAY-1996  (first entry)
DE   Human modified C3 (R1303X).
KW   C3 protein; convertase; Factor I; Factor H; complement;
KW   tumour; infection; therapy.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT   /label= Sig_peptide
FT   Protein         23..667
FT   /note= "C3 beta chain"
FT   Peptide         668..671
FT   /note= "amino acids 668-671 are removed when the
FT   precursor is cleaved into the alpha and
FT   beta chains"
FT   Protein         672..1663
FT   /note= "C3 alpha chain"
FT   Misc_difference 1303
FT   /label= Glu, Gly, Gln
PN   WO9607738-A2.
PD   14-MAR-1996.
PF   08-SEP-1995; G02121.
PR   08-SEP-1994; GB-018147.
PR   04-MAY-1995; GB-009102.
PA   (IMUT-) IMUTRAN LTD.
PI   Farries TC,  Harrison RA;
DR   WPI; 96-171613/17.
PT   Mutant complement pathway protein forming stable C3 convertase  -
PT   for generalised complement depletion or localised complement
PT   activation
PS   Claim 8; Fig 1; 81pp; English.
CC   A modified human C3 protein (R94029) differs from the wild-type
CC   (R94028) by substitution of Arg-1303 by glutamic acid, glycine
CC   or glutamine.  It is obtained by site-directed mutagenesis of
CC   C3-encoding cDNA (T17738).  The modification results in improved
CC   resistance to cleavage by Factor I in comparison to wild-type C3.
```



```
CC   This allows the modified C3 to be used therapeutically to
CC   super-active the complement system or the increase a target's
CC   (e.g. tumour, pathogen or virus-infected cell) sensitivity to
CC   complement-mediated destruction.
SQ   Sequence   1663 AA;

  Query Match            19.4%;  Score 2335;  DB 16;  Length 1663;
  Best Local Similarity  28.3%;  Pred. No. 2.35e-194;
  Matches  488;  Conservative  457;  Mismatches 662;  Indels 115;  Gaps  90;

Db       8 sll-ll-llthlplalg-spmysiitpnilrleseetmvleahdaqgdvpvtvtvhdfpg 64

Qy       2 GLLGILCFLIFLGKTWGQEQTYVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSYPD 61

Db      65 kklvlssektvltpatnhmgnvtftipanrefksekgrnkfvtvqatfgtqvve-kvvlv 123

Qy      62 KKFSYSSGHVHLSSENKFQNSAILTIQP-KQLPGGQNPVSVYVLEVV-SKHFSKSKRMPI 119

Db     124 slqsgylfiqtdktiytpgstvlyriftvnhkllpvgrtvmvnienpegipvkqdslssq 183

Qy     120 TYDNGFLFIHTDKPVYTPDQSVKVRVYSLNDDLKPAKRETVLTFIDPEGSEV--DMVEEI 177

Db     184 nqlgvlpls-wdipelvnmgqwkirayyenspqqvfstefevkeyvlpsfeviveptekf 242

Qy     178 DHIGIISFPDFKIPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEP-E-Y 235

Db     243 yyi-yne-kglevtitarflygkkv-egtafvifgi-qd--geqrislpeslkripiedg 296

Qy     236 NFIGYKNFKNFEITIKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLING 295

Db     297 sgevvlsrkvlldgvqnpraedlvgkslyvsatvilhsgsdmvqaersgipivtspyqih 356

Qy     296 IAQVTFDSETAVKELSYYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLN 355

Db     357 ftktpkyfkpgmpfdlmvfvtnp-d---gs-pay-rvp-vav-q--ge-dtvqsltq-gd 404

Qy     356 LVATPLFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDD 415

Db     405 gvaklsinthpsqkplsitvrtkkqelseaeqatrtmqalpystvgnsnnylhlsvlrte 464

Qy     416 GVASFVLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKA 475

Db     465 lrpgetlnvnfllrmdraheakiryytylimnkgrllkagrqvrepgqdlvvlplsittd 524

Qy     476 LLVGEHLNI-IVTPKS-PYIDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQN 533

Db     525 fipsfrlvayytligasgqrevvadsvwvdvkdscvgslvvksgqsedrqpvpgqqmtlk 584

Qy     534 MVPSSRLLVYYIVTGEQTA-ELVSDSVWLNIEEKCGNQLQVHLSPDADAYS-PGQTVSLN 591

Db     585 iegdhgarvvlvavdkgvfvlnkknkltqskiwdvvekadigctpgsgkdyagvfsdagl 644

Qy     592 MATGMDSWVALAAVDSAVYGVQRGAKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLAGL 651

Db     645 tftsssgqqtaqraelqcpqpaarrrrsvqltekrmdkvg-kypkel-rkccedgmrenp 702

Qy     652 TFLTNANADDSQENDEPCKE-ILRPRRT--L-QKKIEEIAAKYKHSVVKKCCYDGACVNN 707

Db     703 mrfscqrrtrfislgeackkvfldccnyitelrrqharashlglarsnldediiaeeniv 762

Qy     708 DE-TCEQRAARISLGPRCIKAFTECCVVASQLRANISH-KDMQLGRLHMK-TLLPVSKPE 764

Db     763 srsefpeswlwnvedlkeppkngistklmniflkdsittweilavsmsdkkgicvadpfe 822
```

```
Qy     765 IRSYFPESWLWEVH-L-VPRRK----QL-QFALPDSLTTWEIQGIGISNT-GICVADTVK 816

Db     823 vtvmqdffidlrlpssvvrneqveiravlynyrqnqelkvrvellhnpafcslattkrrh 882
               |  | |::: :| |||| ||::::: :|||| :  :    | :     ::|:  :     |
Qy     817 AKVFKDVFLEMNIPYSVVRGEQIQLKGTVYNYRTS-GMQFCVKMSAVEGICTSESPVIDH 875

Db     883 qqtit---ippk--ssls--vpyvivplktglqevevkaavyhhfisdgvrkslkvvpeg 935
               | | :    :   |   :| |   | : ::||  ||::::          |  : : |:|:||||
Qy     876 QGTKSSKCVRQKVEGSSSHLVTFTVLPLEIGLHNINFSLETW--FGKEILVKTLRVVPEG 933

Db     936 irmnktvavrtldperl-gregvqkedip-padlsdqvpdtesetril-lqgtpvaqmte 992
               :: :   :| ||||  : |  : :|| :|    | | || ||    ||| : |  |:::
Qy     934 VKRESYSGV-TLDPRGIYGTISRRKE-FPYRIPL-DLVPKTEI-KRILSVKGLLVGEILS 989

Db     993 davdaerlkhlivtpsgcgeqnmigmtptviavhyldeteqwekf-g--lekrqgaleli 1049
                 :   | :: |    | | :| ::::: |       |||:  ::|: | :   |  :|       :
Qy     990 AVLSQEGINILTHLPKGSAEAELMSVVPVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKL 1049

Db    1050 kkgytqqlafrqpssafaafvkrapstwltayvvkvfslavnliaidsqvlcgavkwlil 1109
               | |       :::| :   :::        ::||||||: ::|::    : :  :     :| :: ||:
Qy    1050 KEGMLSIMSYRNADYSYSVWKGGSASTWLTAFALRVLGQVNKYVEQNQNSICNSLLWLVE 1109

Db    1110 ekqkpdgvfqedapvihqemigglr-nnnekdmalta.vlislqeakdiceeqvnslpgs 1168
               :  |   :| | |::          :  |  |   :    |: :  |||| :|::: | |||    |: :  :
Qy    1110 NYQLDNGSFKENSQYQPIKLQGTLPVEARENSLYLTAFTVIGIRKAFDICP-LVK-IDTA 1167

Db    1169 itkagdfleanymnlqrsytvaiagyala--qmgr--lkg--pllnk-flttakdk-nr- 1219
               : || :|| | :  | ::|:||::|||:  :: :   :::   : |::   |  ::          |
Qy    1168 LIKADNFLLENTLPAQSTFTLAISAYALSLGDKTHPQFRSIVSALKREALVKGNPPIYRF 1227

Db    1220 w-ed-pgkq--l-yn-----veatsyallallqlkdfdfvppvvrwlneqryyggqygst 1269
               | ::     |:   :                  ||:|:||||: | |||:::| ||::||:|:: ||||: ||
Qy    1228 WKDNLQHKDSSVPNTGTARMVETTAYALLTSLNLKDINYVNPVIKWLSEEQRYGGGFYST 1287

Db    1270 qatfmvfqalaqyqkdapdhqelnldvslqlpsxsskithrihwesasllrseetkeneg 1329
               | |:  :::|::|          :  |::|: :         ::       ::   :   | |: |     |:
Qy    1288 QDTINAIEGLTEYSL-LVKQLRLSMDIDVSYKHKGALHNYKMT-DKNFLGRPVEVLLNDD 1345

Db    1330 ftvtae-gkgqgtlsvvtmyhakakdqltcnkfdlkvtikpapetekrpqdakntmilei 1388
               : |::  | | :|: | |: |  :  :   |: | ||:           :   |        : : :
Qy    1346 LIVSTGFGSGLATVHVTTVVHKTSTSEEVCS-FYLKIDTQDIEASHYRGYGNSDYKRIVA 1404

Db    1389 ctryr-g-dqdat-ms--ildismmtgfapdtddlkqlangvdryiskyeldkafsdrnt 1443
               |: |: : :  : ::   |   ::|||: ||:::: :||| | :|||: :: |:: |   :   :
Qy    1405 CASYKPSREESSSGSSHAVMDISLPTGISANEEDLKALVEGVDQLFTDYQI-KD-G--H- 1459

Db    1444 liiyldkvshseddclafkvhqyfnveliqpgavkvyayynleesctrfyhpekedgkln 1503
               :|: |: :: |:   |: |:: : |:| :: |::   || |     :   || ||  :    : |:
Qy    1460 VILQLNSIPSSDFLCVRFRIFELFEVGFLSPATFTVYEYHRPDKQCTMFY-S-TSNIKIQ 1517

Db    1504 klcrdelcrcaeencf-iqksdd-kvtleerldkacepgvdyvyktrlvkvqlsndfdey 1561
               |:|   |:| | :|  :|    |   :: | | : || | : | ||   :  : : | |   |
Qy    1518 KVCEGAACKCVEADCGQMQEELDLTISAETRKQTACKPEIAYAYKVSITSITVENVFVKY 1577

Db    1562 imaieqtiksgsdevqvgqqrtfispikcrealkleekkhylmwglss-df-wgekpnls 1619
               :: :  |:|     :    : |||   : | :|   |     ::||: |    : ::
Qy    1578 KATLLDIYKTGEAVAEKDSEITFIKKVTCTNA-ELVKGRQYLIMGKEALQIKYNFSFRYI 1636

Db    1620 yiigkdtwvehwpeedecqdeenqkqcqdlgaftesmvvfgc 1661
               | :     ||:| || :  |    :       :|  |:| : : ||
Qy    1637 YPLDSLTWIEYWPRDTTCSSCQAF-L-ANLDEFAEDIFLNGC 1676
```

RESULT 6


```
ID   R63223 standard; Protein; 1642 AA.
AC   R63223;
DT   30-JUN-1995  (first entry)
DE   Cobra CVF1 .
KW   Cobra; C3; third component of complement; human; mouse; rat;
KW   X. laevis; pre-pro molecule; beta chain; alpha chain; codon usage;
KW   G+C content; immune response; host defence; tumour; ss.
OS   Naja naja.
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT   /note= "Signal peptide"
FT   Protein         23..1642
FT   Protein         23..649
FT   /note= "Alpha chain"
FT   Protein         733..1032
FT   /note= "Gamma chain"
FT   Protein         1264..1642
FT   /note= "Beta chain"
PN   WO9423024-A.
PD   13-OCT-1994.
PF   07-APR-1994; U03441.
PR   07-APR-1993; US-043747.
PA   (GEOU ) UNIV GEORGETOWN.
PI   Bredehorst R,  Fritzinger DC,   Vogel C;
DR   WPI; 94-333186/41.
DR   N-PSDB; Q77790.
PT   DNA encoding cobra C3, CVF 1 and CVF 2 - which are used in the
PT   treatment of cancer
PS   Claim 3; Fig 2W-2J(10); 155pp; English.
CC   This sequence represents the cobra venom factor 1 (CVF1).  The CVF1
CC   protein is contained within cobra venom and is a structural analog of
CC   cobra C3 (third component of complement).  It is thought that CVF1 and
CC   C3 are derived from different but related genes.  The thioester site
CC   and factor H binding site of C3 and CVF1 are very similar, even though
CC   neither are actually present in the mature CVF1 protein.  Factors H and
CC   I are able to regulate the activity of complement by dissociating C3b,
CC   Bb, and by cleaving C3b, CVF1 is resistant to this regulation.  CVF1
CC   and CVF2 may be used in the treatment of cancer by linking them to
CC   an antibody which recognises a specific tumour cell.  As CVF1 and CVF2
CC   are insensitive to factor H control this method will lead to the
CC   selective destruction of cancer cells.
SQ   Sequence   1642 AA;

  Query Match            18.4%;  Score 2213;  DB 12;  Length 1642;
  Best Local Similarity  29.3%;  Pred. No. 2.66e-183;
  Matches   502;  Conservative  438;  Mismatches 642;  Indels 132;  Gaps 103;

Db       4 malylvaalligfpg-s-shgalytlitpavlrtdteeqilveahgdstpkqldifvhdf 61
             |:|  :  :|| | | : ::   | : :| ::|  : | |:::  | : : :  | :  :
Qy       1 MGLLGILCFLI-FLGKTWGQEQTYVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSY 59

Db      62 prkqktlfqtrvdmnpaggmlvtptieipakevstdsrqnqyvvvqvtgpqvrlekvvll 121
             | |  :    :| :::       :: : | :|:::         || ::| : :     | : :
Qy      60 PDKKFSYSSGHVHLSSENKFQNSAILTIQPKQLPGGQNPVSYVYLEVVSKHFSKSKRMPI 119

Db     122 syqssflfiqtdkgiytpgspvlyrvfsmdhntskmnktvivefqtpegilvs-snsvd- 179
             :|:::||||:||| :||| :|  ||:|:: :      ::  :: |  |||  |     : :|
Qy     120 TYDNGFLFIHTDKPVYTPDQSVKVRVYSLNDDLKPAKRETVLTFIDPEGSEVDMVEEIDH 179

Db     180 lnff-wp-ynlpdlvslqtwrivaky-ehspenytayfdvrkyvlpsfevrlqpsekff- 235
             : :: |  :::|      | | | ||| |  :    ||||:|: |||| | | ::|  :|:
Qy     180 IGIISFPDFKIPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEPEYNFIG 239

Db     236 yidgnenfhvsitarylygeev-egvafvlfgvk--i-ddakksipdsltripiidgdgk 291
```

```
Qy    240 YKNFK-NFEITIKARYFYNKVVTEADVYITFGIREDLKDDQKEMMQTAMQNTMLINGIAQ 298

Db    292 atlkrdt-frs-rfpnlnelvghtlyasvtvmtesgsdmvvteqsgihivaspyqihftk 349
Qy    299 VTFDSETAVKELSYYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLNLVA 358

Db    350 tpkyfkpgmpyeltvyvtnp-d---gs-paa-hvp-v-ve-ea--fh-smgtt-lsdgta 396
Qy    359 TPLFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDDGVA 418

Db    397 klilniplnaqslpitvrtnhgdlprerqatksmtaiayqtqggsgnylhvaitsteikp 456
Qy    419 SFVLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKALLV 478

Db    457 gdnlpvnfnvkgnanslkqikyftylilnkgkifkvgrqprrdgqnlvtmnlhitpdlip 516
Qy    479 GEHLNI-I-VTPKSPYIDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQNMVP 536

Db    517 sfrfvayyqv-gnn--eivadsvwvdvkdtcmgtl-v-vkgdnliqmpgaamkiklegdp 571
Qy    537 SSRLLVYYIVTGEQTAELVSDSVWLNIEEKCGNQLQVHLSPDADAYSPGQTVSLNMATGM 596

Db    572 garvglvavdkavyvlndkykisqakiwdtieksdfgctagsgqnnlgvfedaglaltts 631
Qy    597 DSWVALAAVDSAVYGVQRGAKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLAGLTFLTN 656

Db    632 tnlntkqrsaakcpqpanrrrrssvllldsnaskaaefqdqdlrkccedvmhenpmgytc 691
Qy    657 ANADDSQENDEPCKE-ILRPRRT---LQKKIEEIAAKYKHSVVKKCCYDGACVNNDE-TC 711

Db    692 ekrakyiqegdackaafleccryikgvrdenqreselflarddnedgfiadsdiisrsdf 751
Qy    712 EQRAARISLGPRCIKAFTECCVVASQLRA-NISHKDMQLGRL-HMKTLLPVSKPEIRSYF 769

Db    752 pkswlwltkdlteepnsqgisskt msfylrdsittwv-lavsftptkgicvaepyeirvm 811
Qy    770 PESWLWEV-HLV--PR-R----KQLQFALPDSLTTWEIQGIGISNT-GICVADTVKAKVF 820

Db    812 kvffidlqmpysvvkneqveirailhnyvnedi-y-vr---velly---npafcsastkg 863
Qy    821 KDVFLEMNIPYSVVRGEQIQLKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTKS 880

Db    864 qryrqqfpikalssravpfvivpleqglhdveikasvqealw-sdgv-rkklkvvpegvq 921
Qy    881 SKCVRQ-KVEGSSSHLVTFTVLPLEIGLHNINF--SL-ET-WFGKEILVKTLRVVPEGVK 935

Db    922 ksivtivkldprak-gvggtqlevikarklddrvpdteietkiiiqgdpvaqiiensidg 980
Qy    936 RESYSGVTLDPRGIYGTISRRKE-FPYRIPLDLVPKTEIKRILSVKGLLVGEILSAVLSQ 994

Db    981 sklnhliitpsgcgeqnmirmaapviaty-yldtteqwetlginrrte-a-vnqivt-gy 1036
Qy    995 EGINILTHLPKGSAEAELMSVV-PVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEGM 1053

Db   1037 aqqmvykkadhsyaaftnrassswltayvvkvfamaakmvagisheiicggvrwlilnrq 1096
Qy   1054 LSIMSYRNADYSYSVWKGGSASTWLTAFALRVLGQVNKYVEQ-NQNSICNSLLWLVENYQ 1112

Db   1097 qpdgafkenapvlsgtmqgg--iqgaeeevyltafilvalleskticndyvnsldssikk 1154
Qy   1113 LDNGSFKENSQYQPIKLQGTLPVEARENSLYLTAFTVIGIRKAFDICP-LVK-IDTALIK 1170

Db   1155 atnyllkkyeklqrpyttaltayalaaadqlndd-r--v--lm-aa-stgrd---h-w-e 1202
```



```
Qy   1171 ADNFLLENTLPAQSTFTLAISAYALSLGDKTHPQFRSIVSALKREALVKGNPPIYRFWKD 1230

Db   1203 --ey-na---ht---hniegtsyallallkmkkfdqtgpivrwltdqnfygetygqtqat 1253
Qy   1231 NLQHKDSSVPNTGTARMVETTAYALLTSLNLKDINYVNPVIKWLSEEQRYGGGFYSTQDT 1290

Db   1254 vmafqalaeyeiqmpthkdlnlditielpdrevpiryrinyenallartvetklnqditv 1313
Qy   1291 INAIEGLTEYSL-LVKQLRLSMDIDVSYKHKGALHNYKMT-DKNFLGRPVEVLLNDDLIV 1348

Db   1314 tas-gdgkatmtiltfynaqlqekanvcnkfhlnvsvenihlna-mg-akgalmlki-ct 1369
Qy   1349 STGFGSGLATVHVTTVVH-KTSTSEEVCS-FYLKIDTQDIEASHYRGYGNSDYKRIVACA 1406

Db   1370 ry-lg--evds-tm-tiidismltgflpdaedltrlskgvdryisryevdnnmaqkvavi 1424
Qy   1407 SYKPSREESSSGSSHAVMDISLPTGISANEEDLKALVEGVDQLFTDYQI-KD-GH---VI 1461

Db   1425 iylnkvshsedeclhfkilkhfevgfiqpgsvkvysyynldekctkfyhpdkgtgllnki 1484
Qy   1462 LQLNSIPSSDFLCVRFRIFELFEVGFLSPATFTVYEYHRPDKQCTMFYSTSN-IKI-QKV 1519

Db   1485 cignvcrcagetcsslnhq-e-ridvplqiekacetnvdyvyktkllrieeqdgndiyvm 1542
Qy   1520 CEGAACKCVEADCGQMQEELDLTISAETRKQTACKPEIAYAYKVSITSITVENVFVKYKA 1579

Db   1543 dvlevikqgtdknprakthqyisqrkcqealnlkvnddyliwgsrsdllptk-d-kisyi 1600
Qy   1580 TLLDIYKTGEAVAEKDSEITFIKKVTCTNA-ELVKGRQYLIMG-K-EALQIKYNFSFRYI 1636

Db   1601 it-kn-twierwphedecqe-eef-qklcddfaq 1630
Qy   1637 YPLDSLTWIEYWPRDTTCSSCQAFLANL-DEFAE 1669
```

RESULT 7
ID R63224 standard; Protein; 1333 AA.
AC R63224;
DT 05-JUL-1995 (first entry)
DE Cobra partial CVF2.
KW Cobra; C3; third component of complement; human; mouse; rat;
KW X. laevis; pre-pro molecule; beta chain; alpha chain; codon usage;
KW immune response; host defence; tumour.
OS Naja naja.
FH Key Location/Qualifiers
FT Region 1..332
FT /note= "Partial alpha chain"
FT Region 416..715
FT /note= "Gamma chain"
FT Region 947..1333
FT /note= "Beta chain"
PN WO9423024-A.
PD 13-OCT-1994.
PF 07-APR-1994; U03441.
PR 07-APR-1993; US-043747.
PA (GEOU) UNIV GEORGETOWN.
PI Bredehorst R, Fritzinger DC, Vogel C;
DR WPI; 94-333186/41.
DR N-PSDB; Q77791.
PT DNA encoding cobra C3, CVF 1 and CVF 2 - which are used in the
PT treatment of cancer
PS Claim 5; Fig 2M-2V; 155pp; English.
CC This sequence represents the C-terminal portion of cobra venom factor

```
CC   2 (CVF2).  A full length sequence encoding pre-pro-CVF2 may be
CC   constructed by ligating this sequence to the 3' end of the 1-964 bp
CC   fragment derived from CVF1 (see also Q77790).  CVF2 is active in both
CC   its glycosylated and unglycosylated states.  Factors H and I are able
CC   to regulate the activity of complement by dissociating C3b, Bb, and by
CC   cleaving C3b, CVF2 is resistant to this regulation.  CVF1 and CVF2 may
CC   be used in the treatment of cancer by linking them to an antibody which
CC   recognises a specific tumour cell.  As CVF1 and CVF2 are insensitive to
CC   factor H control this method will lead to the selective destruction of
CC   cancer cells.
SQ   Sequence   1333 AA;

  Query Match             14.0%;  Score 1692;  DB 12;  Length 1333;
  Best Local Similarity   29.1%;  Pred. No. 3.18e-136;
  Matches  398;  Conservative  340;  Mismatches  528;  Indels  103;  Gaps   74;

Db     10 teqsgihivtspyqiyftktpkyfkpgmpyeltvyvtkp-d---gs-paa-hvp-v-vs- 60
          :| :|| | ||| : :  || ::|||:|| : | |  : |   |: |   :   : |:
Qy    340 AEIPGIKYVLSPYKLNLVATPLFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQ 399

Db     61 ea--ih-segtt-lsdgtaklflntpqnaqslpitvrvnkgdlprerqaiksmtatayqt 116
          |:  :  | : | : || | : || |     | : |:|:  ||| | ||  :  | || :
Qy    400 ETSDLDPSKSVTRVDDGVASFVLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSS 459

Db    117 qggsgnylhvaitsteikpgdnlpvnfnvrgnanslnqikyftylilnkgkifkvgrqhr 176
           : |  |: :      :  |::| : : |  ::  :: |  : ||||:||||:  | : :
Qy    460 LSQSYLYIDWTDNHKALLVGEHLNI-I-VTPKSPYIDKITHYNYLILSKGKIIHFGTREK 517

Db    177 gdgenlvtmnlhitpdlipsfrfvayyqv-gnn--eivadsvwvdvkdtcmgtlvvkgat 233
             :  ::|: :| :::|| |:: || | |:   |::|:||||::: : |    | |  :
Qy    518 FSDASYQSINIPVTQNMVPSSRLLVYYIVTGEQTAELVSDSVWLNIEEKCGNQLQVHLSP 577

Db    234 srdnriqmpgaamkiklegdpgawiglvavdkaeyvlndkykisqakiwdtieksdfgct 293
           |      || :: :::     :|::| ||| | | :     | :  :: : :||||:||
Qy    578 DADA-YS-PGQTVSLNMATGMDSWVALAAVDSAVYGVQRGAKKPLERVFQFLEKSDLGCG 635

Db    294 agsgqnnlgvfedaglalttstnlntkqrsaakcpqpanrrrrssvllldsnaskaaqfq 353
          ||:| ||  ||  |||:: |::| :  | :   | :   | ||:    |       || :
Qy    636 AGGGLNNANVFHLAGLTFLTNANADDSQENDEPCKE-ILRPRRT---LQKKIEEIAAKYK 691

Db    354 dqdlrkccedgmhenpmghtcekrekyiqegdackaaflecchyikgiqddnkreselfl 413
           ::||| ||   |     ||| |   |   |   | || |||      :: :  :  :: |
Qy    692 HSVVKKCCYDGACVNN-DETCEQRAARISLGPRCIKAFTECCVVASQLRANISHK-DMQL 749

Db    414 arsdfeddlfgegnitsrsdfpeswlwlmeqlsehpnskgisskivpfylrdsittwell 473
          :|     |: ::    || ||||||| : :|    |   |       :  | | ||:||||:
Qy    750 GRLHMKT-LLPVSKPEIRSYFPESWLWEV-HLV--PRRK----QL-QFALPDSLTTWEIQ 800

Db    474 avglsptkgicvaepyeitvmkdffidlqlpysvvknuqvkiravlynyadkdi-y-vr- 530
          ::|:| | |||||:       | || |::: :||||: ||: ::: :|||       : : |:
Qy    801 GIGISNT-GICVADTVKAKVFKDVFLEMNIPYSVVRGEQIQLKGTVYNYRTSGMQFCVKM 859

Db    531 --vellyspafcsastesqry-r-eqlpikalssravsfvivpleqglhdvevtasvqge 586
            || : ::        :: :   :   :   : : ||: |:| ::||| |||:::     |::
Qy    860 SAVEGICTSESPVIDHQGTKSSKCVRQKVEGSSSHLVTFTVLPLEIGLHNINF--SLETW 917

Db    587 lmsdgvkkklkvvpegewksivtiieldphtk-giggtqvelvkanklndrvpdteietk 645
          :  : : | |:|||||  :    : : |||:    |   : : |:       | | || |||
Qy    918 FGKEILVKTLRVVPEGVKRESYSGVTLDPRGIYGTISRRKEFPYRIPL-DLVPKTEIKRI 976

Db    646 itiqgdpvaqtiensidgsklnhliitpfgcgeqnmirmtapviaty-yldttqqwetlg 704
          ::: |  |:: :    :      :| |    | | :| ::: :   ||:  : ||:| :|: :
Qy    977 LSVKGLLVGEILSAVLSQEGINILTHLPKGSAEAELMSVV-PVFYVFHYLETGNHWNIFH 1035
```



```
Db    705 inrrte-a-vnqimt-gyaqqlvykkadhsyaaftnsassswltayvvkifalaakivkd 761
           :   |   :: :  |    : |::|| ||:    :::|:||||: :::::   | | :
Qy   1036 SDPLIEKQKLKKKLKEGMLSIMSYRNADYSYSVWKGGSASTWLTAFALRVLGQVNKYV-E 1094

Db    762 inheivcggmrwlilnrqrtdgvfrenapvlfgtmqggiqgaepegsl--tafilvalle 819
          |:: :| :: ||: | |   :| |:||:          :|| :      | ||  ||| ::::
Qy   1095 QNQNSICNSLLWLVENYQLDNGSFKENSQYQPIKLQGTLPVEARENSLYLTAFTVIGIRK 1154

Db    820 srsicnayinildssiskatdyllkkyeklqrpyttaltayalaaaerlndd-r--v--l 874
          :  || : ::| |::: || ::|| :       |   :| |::||||: ::: : : |  |  |
Qy   1155 AFDIC-PLVKI-DTALIKADNFLLENTLPAQSTFTLAISAYALSLGDKTHPQFRSIVSAL 1212

Db    875 m-aa-stgrn---r-w-e-----e---pnaht-hniegtsyallallkmkkfveagpvvq 918
          :  |   |      | | :     :    ||: | : :| |:||||: |::| :     ||:
Qy   1213 KEEALVKGNPPIYRFWKDNLQHKDSSVPNTGTARMVETTAYALLTSLNLKDINYVNPVIK 1272

Db    919 wlidqqyyggtyqqtqatvmmfqalaeyeiqmpthkdlnlditielpdrevpiryrinye 978
          || ::| ||| :  || |:  :::|:|| : :  :  |::|| :     :       |:: :
Qy   1273 WLSEEQRYGGGFYSTQDTINAIEGLTEYSL-LVKQLRLSMDIDVSYKHKGALHNYKMT-D 1330

Db    979 nallaqtvetklnedftvsas-gdgkatmtiltvynaqlredanvcnkfhldvsvenvql 1037
          : :|:: ||  ||:|: ||:: | | ||: : || :         :||: | | :  ::::
Qy   1331 KNFLGRPVEVLLNDDLIVSTGFGSGLATVHVTTVVH-KTSTSEEVCS-FYLKIDTQDIEA 1388

Db   1038 nlkeakgakgalklkictry-lg--evds-tm-tiidvsmltgfvpdtedltrlskgvdr 1092
          :   : |       ::  |: |  :  |  | :  :::|:|: ||: :: ||| | |||:
Qy   1389 SHYRGYGNSDYKRIVACASYKPSREESSSGSSHAVMDISLPTGISANEEDLKALVEGVDQ 1448

Db   1093 yismfeinnnmaqkgtviiyldkvshsedeclhfkilkhfevgfiqpgsvkvysyynlde 1152
           :: ::|::  ::    ||: |: :: |:  |::|:|:  |||||: |::  || |   |
Qy   1449 LFTDYQIKD--GH---VILQLNSIPSSDFLCVRFRIFELFEVGFLSPATFTVYEYHRPDK 1503

Db   1153 kctkiyhpdeatgllnkicvgnvcrcaeetcsllnqqknvt-r-qlriqkafdpnvdyvy 1210
           ||::|    :     : |:| |  |:| |  |: : :: :::|     :  | | |  |:: | |
Qy   1504 QCTMFYSTSNIK--IQKVCEGAACKCVEADCGQMQEELDLTISAETRKQTACKPEIAYAY 1561

Db   1211 ktkllrieekdgndiyvmdvlevlkqgtdqnqqvkvrqyvsqrkcqealnlmvnndyliw 1270
           |  :  |   :     |    :|::: | |   :       ::    | :| :|:   :|||
Qy   1562 KVSITSITVENVFVKYKATLLDIYKTGEAVAEKDSEITFIKKVTCTNA-ELVKGRQYLIM 1620

Db   1271 gpss-dl-wpmkdkisylitkntwierwphedkcqe-eef-qklcddfa 1315
           |  : ::      :  | :   |||| ||::  |   : |  :| |:||
Qy   1621 GKEALQIKYNFSFRYIYPLDSLTWIEYWPRDTTCSSCQAFLANL-DEFA 1668

RESULT    8
ID   R11749 standard; Protein; 1484 AA.
AC   R11749;
DT   04-JUN-1991  (first entry)
DE   Human alpha-2 macroglobulin bait region mutant.
KW   Human alpha-2 macroglobulin; growth media; proteinase inhibitor.
OS   Homo sapiens.
FH   Region           701..759
FH   /label= PZP bait region
FT   Key              Location/Qualifiers
PN   WO9103557-A.
PD   21-MAR-1991.
PF   29-AUG-1990; DK0225.
PR   29-AUG-1989; DK-004237.
PR   29-AUG-1989; DK-004235.
PR   29-AUG-1989; DK-004236.
PA   (NOVO ) NOVO NORDISK A/S.
PI   Esper B, Lars S-J;
DR   WPI; 91-102075/14.
```

```
DR   N-PSDB; Q11581.
PT   DNA encoding alpha macro-globulin - used to produce recombinant
PT   protein for use in growth media, proteinase inhibitors and as
PT   carriers
PS   Disclosure; page 47; 78pp; English.
CC   This alpha-2 macroglobulin bait mutant comprises a bait region from
CC   the human pregnancy zone protein (PZP).  It has a proteinase inhib-
CC   itor profile similar to that of the human PZP.  It is useful as an
CC   additive to growth-media and as a carrier in gene- and enzyme rep-
CC   lacement-therapy as well as being a proteinase inhibitor..
CC   See also Q11176.
SQ   Sequence   1484 AA;

  Query Match            5.1%;  Score 611;  DB 2;  Length 1484;
  Best Local Similarity  22.8%;  Pred. No. 4.56e-40;
  Matches  352;  Conservative  408;  Mismatches 653;  Indels 132;  Gaps 114;

Db     11 lvll-llvllptaasvsgkpq-ymvlvpsllhtettekgcvll-sylne-tvtvsa-sle 65

Qy      1 MGLLGILCFLIFLGKTWGQEQTYVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSYP 60

Db     66 svr-gnrslftdleaendvlhcvpfav-pks-ssnee-vmfltvqvkgptqefkkrttvm 121

Qy     61 DKKFSYSSGHVHLSSENKFQNSAILTIQPKQLPGGQNPVSYVYLEV-VSKH-FSKSKRMP 118

Db    122 vknedslvfvqtdksiykpgqtvkfrvvsmdenfhplneliplvyiqdpkgnriaqwqsf 181

Qy    119 ITYDNGFLFIHTDKPVYTPDQSVKVRVYSLNDDLKPAKRETVLTFI-DPEGSEVDMVEEI 177

Db    182 qlegglkqfs-fplssepfqgsykvvvq-kksggrtehp-ftveefvlpkfevqvtvp-k 237

Qy    178 D-HIGIISFPDFKIPSNPRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEPEYN 236

Db    238 iitile-eemnvsvcglytygkpvp-ghvtvsicrkysdasdchgedsq-afcekfsgql 294

Qy    237 FIGYKNFKNFEITIKARYFYNKVVTEADVYITFGIR-EDLKDDQKEMMQTAMQNTMLING 295

Db    295 nshgcfyqqvktkvfql-krkeyemk-lhteaq-iqeegtvveltgrqsseitrtitkls 351

Qy    296 IAQVTFDSETAVKELSYYSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLN 355

Db    352 fvkvdshfrqgipffgqvrlvdgkgvpipnkvifirgneanyysnatt-d-ehglvqfsi 409

Qy    356 LVATPLFLKPGIPYPIKVQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDD 415

Db    410 nttn-vmgtsltvrv-nykdrspcygyqwvseeheeahhtaylvfspsksfvhlepmshe 467

Qy    416 GVASFVLNLPSGVTVLEFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKA 475

Db    468 lpcghtqtvqahyilnggtllglkklsfyylimakggivrtgthgllvkqedmkghfsis 527

Qy    476 LLVGE-H-LNI-IVTPKSPYID-KITHYNYLILSKGKIIHFGTRE---KFSDASYQ-SIN 527

Db    528 ipvksdiapvarlliy-avl--ptgdvigdsakydven-clankvdlsfsp-sqslpash 582

Qy    528 IPVTQNMVPSSRLLVYYIVTGEQTAELVSDSVWLNIEEKC-GNQLQVHLSPDADAYSPGQ 586

Db    583 a-hlrvtaapqsvcalravdqsvllmkpdaelwassvynllpekdlt-gfpgplndqdd- 639

Qy    587 TVSLNMATGMDSWVALAAVDSAVYGVQRGAKKPLERVFQFLEKSDLGCGAGGGLNNANVF 646

Db    640 e-dcin-rhnvyingit-yt-pvsst-nek-dmysflrdrglkaftnskirkp--k-mcp 690

Qy    647 HLAGLTFLTNANADDSQENDEPCKEILRPRRTLQKKIEEIAAK-YKHSVVKKCCYDGACV 705
```



```
Db    691 qlqsvsagavgqgyygag-l-gvverpyvp-ql-g--tynviplnneqssg--pvpet-v 741

Qy    706 NNDETCEQRAARISLGPRCIKAFTECCVVASQLRANISHKDMQLGRLHMKTLLPVSKPEI 765

Db    742 rkyfpetwiwdlvvvnsagvaevgvtvpdtitewkagafclsedaglgisstaslrafqp 801

Qy    766 RSYFPESWLWEVHLVPRR--KQLQFALPDSLTTWEIQGIGIS-NTGICVADTVKAKVFKD 822

Db    802 ffveltmpysvirgeaftlkatvlnylpkcirvsvqleaspaf-la-vpvekeqaphci- 858

Qy    823 VFLEMNIPYSVVRGEQIQLKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTKSSK 882

Db    859 cangrqtvswavtpkslgnvnftvsaealesqelcgtevpsvpehgrkdtvikpllvepe 918

Qy    883 CV--RQKVE-GSSSH-L--VTFTV-L-PLEI-GLHNINFSLETWFG-KEILVKTLRVVPE 932

Db    919 glekettfnsl-lcpsg--gevs--eelslklppnvveesa--ra-sv-svl-gdilgsa 968

Qy    933 GVKRES-YSGVTLDPRGIYGTISRRKEFPYRIPLDLVPKTEIKRILSVKGLLVGEILSAV 991

Db    969 m-qntqnll-qmpygcgeqnmvlfapniyvldylnetqqltpeikska-ig-y-lntgy- 1022

Qy    992 LSQEGINILTHLPKGSAEAELMSVVPVFYVFHYLETGNHWN-IFHSDPLIEKQKLKKKLK 1050

Db   1023 qrqlnykhy-dgsystfgerygrnqgntwltafvlktfaqarayifideahitqaliwls 1081

Qy   1051 ECMLSIMSYRNADYS-YSVWKG-GSASTWLTAFALRVLGQVNKYVEQNQNSICNSLLWLV 1108

Db   1082 qr-qkdngcfrssgsllnnaikggvedevtlsa-yit-ialleipltv-thpvvrnalfc 1137

Qy   1109 ENYQLDNGSFKENSQYQPIKLQGTLPVEARENSLYLTAFTVIGIRKAFDICPLVKIDT-A 1167

Db   1138 lesawktaqegdhgsh-vytkallayafalagnqdkr-kevlkslneeavkkdnsvhwer 1195

Qy   1168 LIKADNFLLENTLPAQSTFTLAISAYALSLGDKTHPQFRSIVSALKREALVKGNPPIY-R 1226

Db   1196 pqkpkapvghfyepqapsae-vemtsyvllayltaqpaptsedltsatnivkwitkqqna 1254

Qy   1227 FWKDNLQHKDSSVPNTGTARMVETTAYALLTSLNLK------DINYVNPVIKWLSEEQRY 1280

Db   1255 qggfsstqhtvvalhalskyga-atftrtgkaaqvtiqssgtfsskfqvdnnnrlllqqv 1313

Qy   1281 GGGFYSTQDTINAIEGLTEYSLLVKQLRLSMDIDVSYKHKGALHN-YKMTDKN-FLGRPV 1338

Db   1314 slpelpgeysmkvt-gegcvylqtslkynilpekeef-pfalgvqtlp-qtcdepk-aht 1369

Qy   1339 EVL-LNDDLIVSTGFGSGLATVHVTTVVHKTSTSEEVCSFYLKIDTQDIEASHYRGYGNS 1397

Db   1370 sf-qislsvsytgsr--sas-nm-aivdvkmvsgfiplkptvkml-ersnhv-srtevss 1422

Qy   1398 DYKRIVACASYKPSREESSSGSSHAVMDISLPTGISANEEDLKALVEGVDQLFTDYQIKD 1457

Db   1423 nhvliyldkv-snqtlslfftvlqdvpvrdlkpaivkvydyyetd 1466

Qy   1458 GHVILQLNSIPSSDFLCVRFRIFELFEVGFLSPATFTVYEYHRPD 1502

RESULT    9
ID    R11334 standard; Protein; 1474 AA.
AC    R11334;
DT    04-JUN-1991  (first entry)
DE    Recombinant human alpha-2 macroglobulin.
KW    Human alpha-2 macroglobulin; growth media; proteinase inhibitor.
```

```
OS   Homo sapiens.
PN   WO9103557-A.
PD   21-MAR-1991.
PF   29-AUG-1990; DK0225.
PR   29-AUG-1989; DK-004237.
PR   29-AUG-1989; DK-004235.
PR   29-AUG-1989; DK-004236.
PA   (NOVO ) NOVO NORDISK A/S.
PI   Esper B, Lars S-J;
DR   WPI; 91-102075/14.
DR   N-PSDB; Q11581.
PT   DNA encoding alpha macro-globulin - used to produce recombinant
PT   protein for use in growth media, proteinase inhibitors and as
PT   carriers
PS   Disclosure; page 34; 78pp; English.
CC   This recombinant human alpha-2 macroglobulin is useful as an add-
CC   itive to growth-media, as a proteinase inhibitor and as a carrier
CC   in gene- and enzyme replacement-therapy.  See also Q11581.
SQ   Sequence   1474 AA;

  Query Match            4.7%;  Score 564;  DB 2;  Length 1474;
  Best Local Similarity  22.9%;  Pred. No. 5.28e-36;
  Matches  350;  Conservative  408;  Mismatches 629;  Indels 141;  Gaps 118;

Db     27 gkpq-ymvlvpsllhtettekgcvll-sylne-tvtvsa-slesvr-gnrslftdleaen 81
Qy     18 GQEQTYVISAPKIFRVGASENIVIQVYGYTEAFDATISIKSYPDKKFSYSSGHVHLSSEN 77

Db     82 dvlhcvafav-pks-ssnee-vmfltvqvkgptqefkkrttvmvknedslvfvqtdksiy 138
Qy     78 KFQNSAILTIQPKQLPGGQNPVSYVYLEV-VSKH-FSKSKRMPITYDNGFLFIHTDKPVY 135

Db    139 kpgqtvkfrvvsmdenfhplneliplvyiqdpkgnriaqwqsfqlegglkqfs-fplsse 197
Qy    136 TPDQSVKVRVYSLNDDLKPAKRETVLTFI-DPEGSEVDMVEEID-HIGIISFPDFKIPSN 193

Db    198 pfqgsykvvvq-kksggrtehp-ftveefvlpkfevqvtvp-kiitile-eemnvsvcgl 253
Qy    194 PRYGMWTIKAKYKEDFSTTGTAYFEVKEYVLPHFSVSIEPEYNFIGYKNFKNFEITIKAR 253

Db    254 ytygkpvp-ghvtvsicrkysdasdchgedsq-afcekfsqqlnshgcfyqqvktkvfql 311
Qy    254 YFYNKVVTEADVYITFGIR-EDLKDDQKEMMQTAMQNTMLINGIAQVTFDSETAVKELSY 312

Db    312 -krkeyemk-lhteaq-iqeegtvveltgrqsseitrtitklsfvkvdshfrqgipffqq 368
Qy    313 YSLEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLNLVATPLFLKPGIPYPIK 372

Db    369 vrlvdgkgvpipnkvifirgneanyysnatt-d-ehglvqfsinttn-vmgtsltvrv-n 424
Qy    373 VQVKDSLDQLVGGVPVILNAQTIDVNQETSDLDPSKSVTRVDDGVASFVLNLPSGVTVLE 432

Db    425 ykdrspcygyqwvseeheeahhtaylvfspsksfvhlepmshelpcghtqtvqahyilng 484
Qy    433 FNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKALLVGE-H-LNI-IVTPK 489

Db    485 gtllglkklsfyylimakggivrtgthgllvkqedmkghfsisipvksdiapvarlliy- 543
Qy    490 SPYID-KITHYNYLILSKGKIIHFGTRE---KFSDASYQ-SINIPVTQNMVPSSRLLVYY 544

Db    544 avl--ptgdvigdsakydven-clankvdlsfsp-sqslpasha-hlrvtaapqsvcalr 598
Qy    545 IVTGEQTAELVSDSVWLNIEEKC-GNQLQVHLSPDADAYSPGQTVSLNMATGMDSWVALA 603
```



```
Db    599 avdqsvllmkpdaelsassvynllpekdlt-gfpqplndqdd-e-dcin-rhnvyingit 654
Qy    604 AVDSAVYGVQRGAKKPLERVFQFLEKSDLGCGAGGGLNNANVFHLAGLTFLTNANADDSQ 663

Db    655 -yt-pvsst-nek-dmysfledmglkaftnskirkp--k-mcpqlqq-yemhgpe-gl-- 703
Qy    664 ENDEPCKEILRPRRTLQKKIEEIAAK-YKHSVVKKCCYDGACVNNDETCEQRAARISLGP 722

Db    704 r-v-gfyesdvmg---rg---h-a-rl--vhvee--phtet-vrkyfaetwiwdlvvvns 748
Qy    723 RCIKAFTECCVVASQLRANISHKDMQLGRLHMKTLLPVSKPEIRSYFPESWLWEVHLVPR 782

Db    749 ayvaevgvtvpdtitewkagafclsedaglgisstaslrafqpffveltmpysvirgeaf 808
Qy    783 R--KQLQFALPDSLTTWEIQGIGIS-NTGICVADTVKAKVFKDVFLEMNIPYSVVRGEQI 839

Db    809 tlkatvlnylpkcirvsvqleaspaf-la-vpvekeqaphci-cangrqtvswavtpksl 865
Qy    840 QLKGTVYNYRTSGMQFCVKMSAVEGICTSESPVIDHQGTKSSKCV--RQKVE-GSSSH-L 895

Db    866 gnvnftvsaealesqelcgtevpsvpehgrkdtvikpllvepeglekettfnsl-lcpsg 924
Qy    896 --VTFTV-L-PLEI-GLHNINFSLETWFG-KEILVKTLRVVPEGVKRES-YSGVTLDPRG 948

Db    925 --gevs--eelslklppnvveesa--ra-sv-svl-gdilgsam-qntqnll-qmpygcg 973
Qy    949 IYGTISRRKEFPYRIPLDLVPKTEIKRILSVKGLLVGEILSAVLSQEGINILTHLPKGSA 1008

Db    974 eqmmvlfapniyvldylnetqqltpeikska-ig-y-lntgy-qrqlnykhy-dgsystf 1028
Qy   1009 EAELMSVVPVFYVFHYLETGNHWN-IFHSDPLIEKQKLKKKLKEGMLSIMSYRNADYS-Y 1066

Db   1029 gerygrnqgntwltafvlktfaqarayifideahitqaliwlsqr-qkdngcfrssgsll 1087
Qy   1067 SVWKG-GSASTWLTAFALRVLGQVNKYVEQNQNSICNSLLWLVENYQLDNGSFKENSQYQ 1125

Db   1088 nnaikggvedevtlsa-yit-ialleilltv-thpvvrnalfclesawktaqegdhgsh- 1143
Qy   1126 PIKLQGTLPVEARENSLYLTAFTVIGIRKAFDICPLVKIDT-ALIKADNFLLENTLPAQS 1184

Db   1144 vytkallayafalagnqdkr-kevlkslneeavkkdnsvhwerpqkpkapvghfyepqap 1202
Qy   1185 TFTLAISAYALSLGDKTHPQFRSIVSALKREALVKGNPPIY-RFWKDNLQHKDSSVPNTG 1243

Db   1203 sae-vemtsyvllayltaqpaptsedltsatnivkwitkqqnaqggfsstqhtvvalhal 1261
Qy   1244 TARMVETTAYALLTSLNLK------DINYVNPVIKWLSEEQRYGGGFYSTQDTINAIEGL 1297

Db   1262 skyga-atftrtgkaaqvtiqssgtfsskfqvdnnnrlllqqvslpelpgeysmkvt-ge 1319
Qy   1298 TEYSLLVKQLRLSMDIDVSYKHKGALHN-YKMTDKN-FLGRPVEVL-LNDDLIVSTGFGS 1354

Db   1320 gcvylqtslkynilpekeef-pfalgvqtlp-qtcdepk-ahtsf-qislsvsytgsr-- 1373
Qy   1355 GLATVHVTTVVHKTSTSEEVCSFYLKIDTQDIEASHYRGYGNSDYKRIVACASYKPSREE 1414

Db   1374 sas-nm-aivdvkmvsgfiplkptvkml-ersnhv-srtevssnhvliyldkv-snqtls 1428
Qy   1415 SSSGSSHAVMDISLPTGISANEEDLKALVEGVDQLFTDYQIKDGHVILQLNSIPSSDFLC 1474

Db   1429 lfftvlqdvpvrdlkpaivkvydyyetd 1456
Qy   1475 VRFRIFELFEVGFLSPATFTVYEYHRPD 1502
```

```
RESULT   10
ID   P71666 standard; Protein; 74 AA.
AC   P71666;
DT   25-APR-1991  (first entry)
DE   Human anaphylatoxin C5a.
KW   activated complement component C5a.
OS   Synthetic.
PN   EP-220864-A.
PD   06-MAY-1987.
PF   10-OCT-1986; 307839.
PR   18-OCT-1985; US-789206.
PR   18-MAR-1986; US-841121.
PA   (PFIZ) PFIZER INC.
PI   Davidow LS, Franke AE, Dezeeuw JR.
DR   WPI; 87-124409/18.
PT   New Yarrowia lipolytica transformants - used fro expression and
PT   secretion of heterologous proteins, esp. prorennin, and human
PT   anaphylatoxin C5a
PS   Example; Fig 9; 45pp; English.
CC   The synthetic gene encoding this protein was constructed from ten
CC   47-mer oligonuleotides. The sequence was chosen to maximise E.coli
CC   and S.cerevisiae preferred codon utilisation and to allow for
CC   several restriction enzyme sites. Direct expression of C5a in
CC   E.coli is posssible by inclusion of an ATG codon in front of the
CC   triplet coding for the first amino acid of the C5a polypeptide.
CC   See also N70213-6, N70218, N71339, N71340, N71343-8.
SQ   Sequence   74 AA;

  Query Match            4.5%;  Score 541;  DB 2;  Length 74;
  Best Local Similarity 100.0%;  Pred. No. 5.04e-34;
  Matches   74;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Db       1 tlqkkieeiaakykhsvvkkccydgacvnndetceqraarislgprcikafteccvvasq 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     678 TLQKKIEEIAAKYKHSVVKKCCYDGACVNNDETCEQRAARISLGPRCIKAFTECCVVASQ 737

Db      61 lranishkdmqlgr 74
           ||||||||||||||
Qy     738 LRANISHKDMQLGR 751


RESULT   11
ID   R75497 standard; Protein; 74 AA.
AC   R75497;
DT   13-NOV-1995  (first entry)
DE   Human C5a protein.
KW   C5a; C5a receptor-antagonist; antiinflammatory.
OS   Synthetic.
PN   WO9516033-A.
PD   15-JUN-1995.
PF   16-NOV-1994; IB0359.
PR   06-DEC-1993; US-162591.
PA   (CIBA ) CIBA GEIGY AG.
PI   Boyar WC,  Galakatos NG,  Peppard JV,  Van Oostrum J;
DR   WPI; 95-224319/29.
DR   N-PSDB; Q92518.
PT   C5a receptor antagonists having no agonist activity - are used in
PT   compsns. to treat C5a-mediated diseases and inflammatory conditions
PS   Disclosure; Page 36-37; 65pp; English.
CC   C5a encoded by a synthetic gene is given in R75497.  Analogs of C5a,
CC   obtd. by mutagenesis of the C-terminal region, exhibit excellent
CC   antagonistic properties and substantially no agonist activity.
SQ   Sequence   74 AA;
```



```
  Query Match            4.5%;  Score 541;  DB 13;  Length 74;
  Best Local Similarity 100.0%;  Pred. No. 5.04e-34;
  Matches   74;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Db       1 tlqkkieeiaakykhsvvkkccydgacvnndetceqraarislgprcikafteccvvasq 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     678 TLQKKIEEIAAKYKHSVVKKCCYDGACVNNDETCEQRAARISLGPRCIKAFTECCVVASQ 737

Db      61 lranishkdmqlgr 74
           ||||||||||||||
Qy     738 LRANISHKDMQLGR 751


RESULT   12
ID   R71933 standard; protein; 403 AA.
AC   R71933;
DT   21-SEP-1995  (first entry)
DE   Cobra venom factor C-terminal region.
KW   Cobra venom factor; CVF; cytostatic; tumor.
OS   Naja naja.
FH   Key              Location/Qualifiers
FT   Modified_site   107
FT   /label= N-glycosylation_site
FT   Disulfide_bond  253
FT   /note= "interchain disulfide linkage to CVF gamma
FT   chain"
PN   WO9507299-A.
PD   16-MAR-1995.
PF   12-SEP-1994; U09770.
PR   10-SEP-1993; US-118674.
PA   (GEOU ) UNIV GEORGETOWN.
PI   Bredehorst R,  Grunwald T,  Ollert MW,  Vogel C,  Ziegelmuller P;
DR   WPI; 95-123385/16.
PT   New cobra venom factor deriv. - produced by protease digestion,
PT   used for activating complement partic. in killing tumour cells
PS   Disclosure; Fig.9; 36pp; English.
CC   New cobra venom factor (CVF) derivs. are obtained by chymotrypsin
CC   or trypsin digestion of the C-terminal region of CVF beta-chain
CC   (given in R71933).  The CVF derivs. retain complement activating
CC   activity but are less prone to neutralization by the immune
CC   response than natural CVF.
SQ   Sequence   403 AA;

  Query Match            4.5%;  Score 540;  DB 13;  Length 403;
  Best Local Similarity  27.7%;  Pred. No. 6.15e-34;
  Matches  112;  Conservative  105;  Mismatches 156;  Indels  32;  Gaps  27;

Db       4 ygetygqtqatvmafqalaeyeiqmpthkdlnlditielpdrevpiryrinyenallart 63
           ||  :  || |: |:::|:|| : :  :  |::|| :     :     |::  :: :|:|
Qy    1280 YGGGFYSTQDTINAIEGLTEYSL-LVKQLRLSMDIDVSYKHKGALHNYKMT-DKNFLGRP 1337

Db      64 vetklnqditvtas-gdqkatmtiltfynaqlqeranvcnrphlnvsvenihlna-hg-a 120
           ||  ||:|: |::: | | ||: : |  :       :||:   |::  ::|  :  :| :
Qy    1338 VEVLLNDDLIVSTGFGSGLATVHVTTVVH-KTSTSEEVCSF-YLKIDTQDIEASHYRGYG 1395

Db     121 kgalmlki-ctry-lg--evds--thtiidismltgflpdaedltrlsrgvdryisryev 174
           ::  :  : |: |  :  |  |  :|:::|||: ||: :: |||  |  |||: :: |::
Qy    1396 NSDYKRIVACASYKPSREESSSGSSHAVMDISLPTGISANEEDLKALVEGVDQLFTDYQI 1455

Db     175 dnnmaqkvaviiylnkvshsedeclhfkilkhfevgfiqpgsvkvysyynlderctkfyh 234
            :: ::   ||: || :: |:  |::|:|:  |||||: |::  || |    | :||:|| 
Qy    1456 -KD-GH---VILQLNSIPSSDFLCVRFRIFELFEVGFLSPATFTVYEYHRPDKQCTMFYS 1510
```

```
RESULT  15
ID   R51948 standard; protein; 344 AA.
AC   R51948;
DT   08-NOV-1994  (first entry)
DE   Phospholipase A2 inhibitor - C3dg (Rat).
KW   Phospholipase A2; Inhibitor; Rat C3dg; production;
KW   organic solvent; acid solvent; Allergy; Inflammation.
OS   Rattus sp.
PN   J06065293-A.
PD   08-MAR-1994.
PF   21-AUG-1992; 222693.
PR   21-AUG-1992; JP-222693.
PA   (TEIJ ) TEIJIN LTD.
DR   WPI; 94-121221/15.
PT   Prodn. of 14kDaII type phospholipase A2 inhibitor - by treating
PT   specified proteins with organic and/or acidic solvent, to obtain
PT   enzyme useful to treat inflammation and allergy
PS   Claim 1; Page 5-6; 8pp; Japanese.
CC   A 14kDaII type phospholipase A2 inhibitor, (C3dg) is produced by
CC   treating the polypeptide R51948 with an organic and/or acid solvent.
CC   Rat C3dg can be mass-produced by genetic engineering for use in
CC   the treatment of inflammation and allergic diseases.
SQ   Sequence   344 AA;

  Query Match            3.6%;  Score 429;  DB 10;  Length 344;
  Best Local Similarity  28.5%;  Pred. No. 1.78e-24;
  Matches  101;  Conservative   98;  Mismatches 127;  Indels  28;  Gaps  17;

Db     10 dqvpdtdsetril-lqgtpvaqmaedavdgerlkhlivtpsgcgeqnmigmtptviavhy 68
           | || |:   ||| : |  |:::     :  | :: |   | | :| ::::: |     ||
Qy    966 DLVPKTEI-KRILSVKGLLVGEILSAVLSQEGINILTHLPKGSAEAELMSVVPVFYVFHY 1024

Db     69 ldqteqwekf-g--lekrqealelikkgytqqlafkqpssayaafnnrppstwltayvvk 125
           |:  ::|: | :  |  :|     :| |     :::: :  :|:  :  ::||||||: ::
Qy   1025 LETGNHWNIFHSDPLIEKQKLKKKLKEGMLSIMSYRNADYSYSVWKGGSASTWLTAFALR 1084

Db    126 vfslaanliaidsqvlcgavkwlilekqkpdgvfqedgpvihqemiggfr-ntkeadvsl 184
           |::   : :  :   :| :: ||: : |  :| | |::       : | :  :::|  : |
Qy   1085 VLGQVNKYVEQNQNSICNSLLWLVENYQLDNGSFKENSQYQPIKLQGTLPVEARENSLYL 1144

Db    185 tafvlialqeardicegqvnslpgsinkagetleasylnlqrpytvaiagyalalmnkle 244
           ||| :|::: | |||    |: :  :: || : |   : |   | :|:||::|||:| :|
Qy   1145 TAFTVIGIRKAFDICP-LVK-IDTALIKADNFLLENTLPAQSTFTLAISAYALSLGDKTH 1202

Db    245 epy--ltkflnt-a--k-drn--r-weepgqq----l-yn-----veatsyallalllk 285
             :  :   |:  |  | :    | | :  |:      :         ||:|:||||: | ||
Qy   1203 PQFRSIVSALKREALVKGNPPIYRFWKDNLQHKDSSVPNTGTARMVETTAYALLTSLNLK 1262

Db    286 dfdsvppvvrwlneqryygggygstqatfmvfqalaqyqtdvpdhkdlnmdvsl 339
           |:: | ||::||:|:: ||||: ||| |:  :::|::|    | : : |:||: :
Qy   1263 DINYVNPVIKWLSEEQRYGGGFYSTQDTINAIEGLTEYSLLVKQLR-LSMDIDV 1315

Search completed: Wed Jan 28 12:13:50 1998
Job time : 66 secs.
```

```
Db     235 pdrctcllnkicignvcrcagetcsslnhq-e-ridvplcieracetnvdyvyktkllri 292
               : |:| |  |:|    |: :  : :  |    : : ||  :: | ||  :  |
Qy    1511 TSNIK--IQKVCEGAACKCVEADCGQMQEELDLTISAETRKQTACKPEIAYAYKVSITSI 1568

Db     293 eeqdcndiyvhdvlevikqgtdknrrakthqyisqrkcqealnlkvnddyliwgsrsdll 352
               ::    |   :|:: | |     :      :|     | :| :|    :||| | : : |
Qy    1569 TVENVFVKYKATLLDIYKTGEAVAEKDSEITFIKKVTCTNA-ELVKGRQYLIMG-K-EAL 1625

Db     353 ptk-d-kisyiit-kn-twierwphedecqe-eef-qklcddfaq 391
               | :  : ||    : |||| ||::  |    :  |  :| |:||:
Qy    1626 QIKYNFSFRYIYPLDSLTWIEYWPRDTTCSSCQAFLANL-DEFAE 1669
```

RESULT 13
ID R21775 standard; Protein; 344 AA.
AC R21775;
DT 25-JUN-1992 (first entry)
DE Phospholipase A2 inhibitory protein.
KW Inflamed tissue; antihistamine; allergic reactions.
OS Rattus rattus.
PN WO9202619-A.
PD 20-FEB-1992.
PF 03-AUG-1991; J01040.
PR 03-AUG-1990; JP-205164.
PR 19-DEC-1990; JP-411594.
PA (TEIJ) TEIJIN KK.
PI Suwa Y, Imaizumi A, Okada M, Suzuki Y, Kudo I, Inoue K;
PI Azumac, Murakami M;
DR WPI; 92-080073/10.
DR N-PSDB; Q21650.
PT Plasmid encoding phospholipase A2 inhibitory protein - derived
PT from human or rat inflamed region, used as antihistamine agents
PT for treating allergic disorders
PS Disclosure; Page 31; 52pp; Japanese.
CC A rat liver cDNA lambda gt11 library obtd. from inflammed tissue
CC was screened using mouse C3cDNA-derived plasmid pFC4/5.4 (cleaved
CC with HindIII/StuI) as probe. The rat C3cDNA clone pTC3/11 obtd.
CC was cleaved with EcoRI into 2.0 kb and 0.1 kb fragments which
CC were combined with vector pTV119N to give ppTC3/2.1. The plasmid
CC was used to transform E. coli and the transformant found to express
CC a phospholipase inhibitory protein of Mw. 39 kD (SEQ ID NO 1). The
CC protein can be used as an anti-histamine agent which can be used
CC to treat allergic reactions.
CC See also R21776.
SQ Sequence 344 AA;

```
  Query Match             3.7%;  Score 450;  DB 4;  Length 344;
  Best Local Similarity   28.8%;  Pred. No. 2.98e-26;
  Matches  102;  Conservative   99;  Mismatches 125;  Indels  28;  Gaps  17;

Db      10 dqvpdtdsetril-lqgtpvaqmaedavdgerlkhlivtpsgcgeqnmigmtptviavhy 68
           | || |:   ||| : |  |:::     :  | :: |    | | :| ::::: |     ||
Qy     966 DLVPKTEI-KRILSVKGLLVGEILSAVLSQEGINILTHLPKGSAEAELMSVVPFVYVFHY 1024

Db      69 ldqteqwekf-g--lekrqealelikkgytqqlafkqpssayaafnnrppstwltayvvk 125
           |:  ::|: | :   |  :|      :| |     :::: :  :|:  :  ::||||||: ::
Qy    1025 LETGNHWNIFHSDPLIEKQKLKKKLKEGMLSIMSYRNADYSYSVWKGGSASTWLTAFALR 1084

Db     126 vfslaanliaidsqvlcgavkwlilekqkpdgvfqedgpvihqemiggfr-ntkeadvsl 184
           |::    :  :  :   :| :: ||: :  |   :| | |::         : | :  :::| : |
Qy    1085 VLGQVNKYVEQNQNSICNSLLWLVENYQLDNGSFKENSQYQPIKLQGTLPVEARENSLYL 1144

Db     185 tafvlialqeardicegqvnslpgsinkageyleasylnlqrpytvaiagyalalmnkle 244
           ||| :|::: | |||    |: :  :: || ::|  : |  | :|:||::|||:| :|
```



```
Qy    1145 TAFTVIGIRKAFDICP-LVK-IDTALIKADNFLLENTLPAQSTFTLAISAYALSLGDKTH 1202

Db     245 epy--ltkflnt-a--k-dpn--r-weepgqq----l-yn-----veatsyallalllk 285
            :  :   |:  |  | :|   | | :  |:    :         ||:|:||||: | ||
Qy    1203 PQFRSIVSALKREALVKGNPPIYRFWKDNLQHKDSSVPNTGTARMVETTAYALLTSLNLK 1262

Db     286 dfdsvppvvrwlneqryyggqygstqatfmvfqalaqyqtdvpdhkdlnmdvsl 339
           |:: | ||::||:|:: ||||: ||| |:  :::|::|    | : : |:||: :
Qy    1263 DINYVNPVIKWLSEEQRYGGGFYSTQDTINAIEGLTEYSLLVKQLR-LSMDIDV 1315
```

RESULT 14
ID R10899 standard; Protein; 344 AA.
AC R10899;
DT 07-MAY-1991 (first entry)
DE Rat phospholipase A2 inhibitor protein.
KW Inflammatory disease.
OS Rattus rattus.
PN WO9101999-A.
PD 21-FEB-1991.
PF 03-AUG-1990; J00996.
PR 03-AUG-1989; JP-200246.
PR 05-APR-1990; JP-089085.
PA (TEIJ) TEIJIN KK.
PI Suwa Y, Imaizumi A, Okada M, Kudo I, Inoue K, Suzuki Y;
DR WPI; 91-073488/10.
DR N-PSDB; Q10773.
PT Phospholipase A2 inhibitor protein - obtd. by enzymic treatment
PT of mammalian serum.
PS Claim 1; Fig 1; 68pp; English.
CC The protein, obtd. from rat serum or expressed from the gene, is
CC used in the treatment and diagnosis of inflammatory diseases.
CC See also R10900.
SQ Sequence 344 AA;

```
  Query Match             3.6%;  Score 436;  DB 3;  Length 344;
  Best Local Similarity   28.2%;  Pred. No. 4.57e-25;
  Matches  100;  Conservative   99;  Mismatches 127;  Indels  28;  Gaps  17;

Db      10 dqvpdtdsetril-lqgtpvaqmadeavdgerlkhlivtpsgcgeqnmigmtptviavhy 68
           | || |:   ||| : |  |:::     :  | :: |    | | :| ::::: |     ||
Qy     966 DLVPKTEI-KRILSVKGLLVGEILSAVLSQEGINILTHLPKGSAEAELMSVVPFVYVFHY 1024

Db      69 ldqteqwekf-g--lekrqealelikkgytqqlafkqpssayaafnnrppstwltayvvk 125
           |:  ::|: | :   |  :|      :| |     :::: :  :|:  :  ::||||||: ::
Qy    1025 LETGNHWNIFHSDPLIEKQKLKKKLKEGMLSIMSYRNADYSYSVWKGGSASTWLTAFALR 1084

Db     126 vfslaanliaidsqvlcgavkwlilekqkpdgvfqedgpvihqemiggfr-ntkeadvsl 184
           |::    :  :  :   :| :: ||: :  |   :| | |::         : | :  :::| : |
Qy    1085 VLGQVNKYVEQNQNSICNSLLWLVENYQLDNGSFKENSQYQPIKLQGTLPVEARENSLYL 1144

Db     185 tafvlialqeardicegqvnslpgsinkageyleasylnlqrpytvaiagyalalmnkle 244
           ||| :|::: | |||    |: :  :: || ::|  : |  | :|:||::|||:| :|
Qy    1145 TAFTVIGIRKAFDICP-LVK-IDTALIKADNFLLENTLPAQSTFTLAISAYALSLGDKTH 1202

Db     245 epy--ltkflnt-a--k-drn--r-weepgqq----l-yn-----veatsyallalllk 285
            :  :   |:  |  | :    | | :  |:    :         ||:|:||||: | ||
Qy    1203 PQFRSIVSALKREALVKGNPPIYRFWKDNLQHKDSSVPNTGTARMVETTAYALLTSLNLK 1262

Db     286 gfdsvppvvrwlneqryyggqygstqatfmvfqalaqyqtdvpdhkdlnmdvsl 339
           :: | ||::||:|:: ||||: ||| |:  :::|::|    | : : |:||: :
Qy    1263 DINYVNPVIKWLSEEQRYGGGFYSTQDTINAIEGLTEYSLLVKQLR-LSMDIDV 1315
```